WORKING PAPERS IN ECONOMICS

No 383

Testing for a Unit Root in a Random Coefficient Panel Data Model

Joakim Westerlund and Rolf Larsson

September 2009

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)
TESTING FOR A UNIT ROOT IN A RANDOM COEFFICIENT PANEL DATA MODEL*

Joakim Westerlund†  Rolf Larsson
University of Gothenburg  Uppsala University
Sweden  Sweden

September 25, 2009

Abstract

This paper proposes a new unit root test in the context of a random autoregressive coefficient panel data model, in which the null of a unit root corresponds to the joint restriction that the autoregressive coefficient has unit mean and zero variance. The asymptotic distribution of the test statistic is derived and simulation results are provided to suggest that it performs very well in small samples.

JEL Classification: C13; C33.

Keywords: Panel unit root test; Random coefficient autoregressive model.

1 Introduction

Consider the panel data variable $y_{it}$, observable for $t = 1, \ldots, T$ time series and $i = 1, \ldots, N$ cross-sectional units. The analysis of such variables has been a growing field of econometric research in recent years, with a majority of the work focusing on the issue of unit root testing, see Breitung and Pesaran (2008) for a recent review. The main reason for this being the well-known power problem of univariate tests in cases when $T$ is small, and the potential

*Previous versions of this paper were presented at the 15th International Conference on Panel Data in Bonn and at a seminar at the University of Gothenburg. The author would like to thank conference and seminar participants, and in particular Jushan Bai, Steven Bond, Jörg Breitung, Dick Durevall, Lennart Flood and Hans Christian Kongsted for many valuable comments and suggestions. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant numbers P2005–0117:1 and W2006–0068:1 is gratefully acknowledged.

†Corresponding author: Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.
gain that can be made by pooling across a cross-section of similar units. The most common approach, pioneered by Levin et al. (2002), is to assume that $y_{it}$ admits to a first-order autoregressive representation with a common slope coefficient,

$$y_{it} = \rho y_{i,t-1} + u_{it},$$

where $u_{it}$ is a stationary disturbance term with zero mean. A pooled least squares $t$-statistic is then computed, and the null hypothesis that $\rho = 1$ is tested against the alternative that $|\rho| < 1$.

The major limitation of this approach is that $\rho$ is restricted to be the same for all units. The null makes sense, but the alternative is too strong to be held in any interesting empirical cases. For example, when testing for price convergence, one can formulate the null as implying that none of the regions under study converges. But it does not make any sense to assume that all the regions will converge at the same rate if they do converge.

Im et al. (2003) relax the assumption of a common autoregressive coefficient under the alternative. The idea is very simple. Take the above model and substitute $\rho_i$ for $\rho$, which in the usual formulation where $\rho_i$ is fixed results in $N$ separate autoregressive models, one for each unit. Thus, instead of looking at a single pooled $t$-statistic, we now look at $N$ individual $t$-statistics, which can be combined for example by taking the average. The resulting average statistic tests the null that $\rho_i = \rho = 1$ for all $i$ against the alternative that $|\rho_i| < 1$ for a positive fraction of $N$.

But this is basically the same as saying that the null should be rejected if at least one of the individual tests end up in a rejection at the appropriate significance level, which brings us back to the original problem, namely that $T$ has to be large. But if $T$ is large enough for valid inference at the individual level, then there is hardly any point in pooling. This leaves us with an intricate dilemma. On the one hand, we would like to exploit the additional power that becomes available when we pool, and when we do this we would like to allow for some heterogeneity in $\rho_i$. On the other hand, this allowance requires $T$ to be large, in which case we can just as well go back to doing unit-by-unit inference.

The appropriate response here depends on the relative size of $N$ and $T$. But if only $N$ is large enough, then it should be possible to devise powerful tests that are informative in an average sense, even if $T$ is small. This leads naturally to the consideration of a random
specification for $\rho_i$. In particular, suppose that

$$\rho_i = 1 + c_i,$$

where $c_i$ is an independently distributed random variable with mean $\mu_c$ and variance $\omega_c^2$. Then the null of a unit root corresponds to the joint restriction that $\mu_c = \omega_c^2 = 0$, while the alternative is that $\mu_c \neq 0$ or $\omega_c^2 > 0$, or both.

This random specification of $c_i$ has many advantages in comparison to the traditional fixed specification. Firstly, working with incompletely specified models inevitably leads to a loss of efficiency. The random specification reduces the number of parameters that need to be estimated, and is therefore expected to lead to more powerful tests. Secondly, the random specification is more general, because fixed coefficients are special random variables. Whether something is random or not should be decided by considering what would happen if we were to replicate the experiment. Is it realistic to assume that $c_i$ stays the same under replication? If not, then the random specification is more appropriate. Thirdly, by considering not only the mean of $c_i$ but also the variance, random coefficient tests account for more information, and are therefore expected to be more powerful. Fourthly, the alternative hypothesis does not rule out the possibility that some of the units may be explosive.

Taking this random coefficient model as our starting point, the goal of this paper is to design a procedure to test the null hypothesis that $\mu_c = \omega_c^2 = 0$, which has not received much attention in the previous literature. In fact, the only attempt that we are aware of is that of Ng (2008), who uses a random coefficient model as a basis for proposing an estimator of the fraction of units with a unit root. However, this procedure does not exploit the fact that under the null hypothesis the variance of $c_i$ is zero, which makes it suboptimal from a power point of view. It is also rather restrictive in nature, and cannot be easily generalized to accommodate for example high-order serial correlation.

Our testing methodology is rooted in the Lagrange multiplier principle, and can be seen as a generalization of the recent time series work of Distaso (2008) and Ling (2004), who consider the problem of testing for a unit root when the autoregressive coefficient is time-varying. It is also very similar to the seminal approach of Schmidt and Phillips (1992), from which it inherits many of its distinctive features. The test is for example based on a very convenient detrending procedure that imposes the null hypothesis, and if a linear trend is included the test statistic is asymptotically invariant with respect to the presence of a level
break. It is also very straightforward and easy to implement.

The asymptotic analysis reveals that the Lagrange multiplier test statistic has a limiting chi-squared distribution that is free of nuisance parameters under the null hypothesis. We also study the limiting behavior of the statistic under local alternative hypotheses. We show that in the case of either a constant that may by heterogeneous across units, or a constant and trend that are homogenous the test has power against alternatives that shrink towards the unit root at rate $\frac{1}{\sqrt{NT}}$. However, we also show that in the presence of a heterogeneous trend the test does not have any power in such neighborhoods, which is a reflection of the so-called incidental trends problem.

A small simulation study is also undertaken to evaluate the small-sample properties of the test, and the results show that the asymptotic properties are borne out well, even in very small samples.

The rest of the paper is organized as follows. Section 2 introduces the model, while Section 3 derives the Lagrange multiplier test statistic and its asymptotic properties, which are evaluated using both simulated and real data in Sections 4 and 5, respectively. Section 6 concludes. Proofs and derivations of important results are provided in the appendix.

A word on notation. The symbols $\to_w$ and $\to_p$ will be used to signify weak convergence and convergence in probability, respectively. As usual, $y_T = O_p(T^r)$ will be used to signify that $y_T$ is at most order $T^r$ in probability, while $y_T = o_p(T^r)$ will be used in case $y_T$ is of smaller order in probability than $T^r$.\footnote{If $y_T$ is deterministic, then $O_p(T^r)$ and $o_p(T^r)$ are replaced by $O(T^r)$ and $o(T^r)$, respectively.} In the case of a double indexed sequence $y_{NT}$, $T$, $N \to \infty$ will be used to signify that the limit has been taken while passing both indices to infinity jointly. Restrictions, if any, on the relative expansion rate of $T$ and $N$ will be specified separately.

## 2 Model and assumptions

The data generating process of $y_{it}$ is given by

$$y_{it} = d_{it} + z_{it},$$

where $d_{it}$ is the deterministic part of $y_{it}$, while $z_{it}$ is the stochastic part. The typical elements of $d_{it}$ include a constant and a linear time trend, and this is also the specification considered here. Specifically, using $p$ to denote the lag length, then $d_{it} = \alpha_i + \beta_i(t - p)$, which nests two
models. In model 1, there is no trend, while in model 2, there is both an intercept and trend. The parameters $\alpha_i$ and $\beta_i$ can be either known or unknown to be estimated along with the other parameters of the model.

The stochastic part is assumed to evolve according to a first-order autoregressive process,

$$z_{it} = \rho_i z_{i,t-1} + u_{it}, \quad (2)$$

or equivalently,

$$\Delta z_{it} = \epsilon_i z_{i,t-1} + u_{it}$$

with the error $u_{it}$ following a stationary and invertible autoregressive process of order $p$,

$$\phi_i(L)u_{it} = \epsilon_{it}, \quad (3)$$

where $\phi_i(L) = 1 - \sum_{j=1}^{p} \phi_{ji}L^j$ is a polynomial in the lag operator $L$ and $\epsilon_{it}$ is an error term that satisfies the following assumptions.

**Assumption 1.**

(a) $\epsilon_{it}$ is independent across both $i$ and $t$ with mean zero, variance $\sigma_i^2 < \infty$ and $E(\epsilon_{it}^3) = 0,$

(b) $\frac{1}{N} \sum_{i=1}^{N} \kappa_i \to \kappa < \infty,$ where $\kappa_i = E(\epsilon_{it}^4)/\sigma_i^4,$

(c) $\alpha_i$, $\beta_i$ and $\phi_i(L)$ are non-random with the roots of $\phi_i(L)$ falling outside the unit circle,

(d) $z_{i0}, ..., z_{ip}$ are $O_p(1).$

**Assumption 2.** $\epsilon_{it}$ is normally distributed.

The assumed independence across $i$ is restrictive but is made here in order to make the analysis of $\rho_i$ more manageable. Some possibilities for how to relax this condition are discussed in Section 3. Normality is also not necessary. More precisely, while needed for deriving the true Lagrange multiplier test statistic, normality is not needed when deriving its asymptotic distribution. The following assumptions are more important in that regard.

**Assumption 3.**

(a) $c_i$ is independent across $i$ with mean $\mu_c$ and variance $\omega_c^2,$
(b) $c_i$ and $e_{it}$ are mutually independent.

**Assumption 4.** $\frac{N}{T} \to 0$ as $N, T \to \infty$.

The requirement that the mean and variance of $c_i$ are equal across $i$ is made for convenience, and can be relaxed as long as the cross-sectional averages of these moments have limits such as $\mu_c$ and $\omega_c^2$, respectively. However, the assumption that $c_i$ and $e_{it}$ are independent is crucial. Assumption 4 is standard when testing for unit roots in panels. The reason is the assumed heterogeneity in $\alpha_i, \beta_i, \phi_i(L)$ and $\sigma_i^2$, whose elimination induces an estimation error in $T$, which is then aggravated when pooling across $N$. The condition that $\frac{N}{T} \to 0$ prevents the estimation from having a dominating effect, see Section 3 for a more detailed discussion and for some results when it fails.

Having laid out the assumptions we now continue to discuss the hypothesis of interest. In the conventional setup when $c_i$ is fixed the null hypothesis of a unit root is formulated as that $\rho_i = 0$ for all $i$, while the alternative hypothesis is usually formulated as in Im et al. (2003). That is, it is assumed that $c_i < 0$ for a significant fraction of $N$, implying that although some of the units may be non-stationary most of them are stationary.

When $c_i$ is random, this formulation changes. The null of a unit root now becomes

$$H_0 : \rho_i = 0 \quad \text{almost surely},$$

which can be written in an equivalent fashion as

$$H_0 : \mu_c = \omega_c^2 = 0.$$

A violation of this null occurs if $\mu_c \neq 0$ or $\omega_c^2 > 0$, or both, implying that while some units may be non-stationary, the probability of this happening is very small. It also implies that there are not just stationary and non-stationary units, but also explosive units, which seems like a relevant scenario in most applications, especially in financial economics, where data tend to exhibit explosive behavior.\(^2\) Explosive behavior is also more likely if $N$ is large, which obviously increases the probability of extreme events regardless of the application considered. There is also the question to what extent researchers can work with regular unit root tests without prior knowledge of the location of the roots.

\(^2\)In Section 5 we consider as an example the housing market of the United States, which has recently experienced a spectacular rise in prices. Periods of hyperinflation and stock markets with rational bubbles are other examples of applications with possibly explosive data, see for example Nielsen (2008) and Phillips et al. (2009).
In any case, with such a formulation of the alternative hypothesis, we only learn whether the test is consistent and if so at what rate. Therefore, to be able to evaluate the power analytically, in this paper we consider an alternative in which \( \rho_i \) is local-to-unity as \( N, T \to \infty \). In particular, the following formulation is adopted:

\[
H_1 : \rho_i = 1 + \frac{c_i}{\sqrt{NT}},
\]

where \( c_i \) again satisfies Assumption 3. This corresponds to an autoregressive coefficient that approaches one with increasing values of \( N \) and \( T \). If \( c_i < 0 \), then \( \rho_i \) approaches one from below and so \( y_{it} \) is locally stationary, whereas if \( c_i > 0 \), then \( \rho_i \) approaches one from above and so \( y_{it} \) is locally explosive. In the limit as \( N, T \to \infty \) we see that \( \rho_i \to 0 \), and hence the distribution of \( \rho_i \) collapses with the mean going to one and the variance going to zero.

The rate of shrinking is given by \( \frac{1}{\sqrt{NT}} \). Coincidentally, this is also the rate of consistency of the pooled least squares estimator of \( \rho_i \) under the null, which is going to turn out to form the basis of our test statistic. Being an estimate of the slope of the mean function, it is logical to expect that the main effect of the local-to-unity specification of \( \rho_i \) is to induce via \( \mu_c \) a non-centrality of the asymptotic distribution of the test statistic.

3 The test procedure

In this section, we first consider the true Lagrange multiplier test statistic, which is based on the assumption that the parameters of the model are all known. We then show how this analysis extends to the more realistic case when the parameters are unknown. Finally, we discuss some generalizations.

3.1 The true Lagrange multiplier test statistic

Define \( w_{it} = \phi_i(L)(y_{it} - d_{it}) \), which in the model with a trend can be written as

\[
w_{it} = \phi_i(L)(y_{it} - \alpha_i - \beta_i(t - p)) = y_{it} - \Phi_i'y_{it} - \mu_i - \beta_i\phi_i(L)(t - p),
\]

whose first difference is given by

\[
\Delta w_{it} = \phi_i(L)(\Delta y_{it} - \beta_i) = \Delta y_{it} - \Phi_i'\Delta y_{it} - \lambda_i,
\]

where \( \mu_i = \phi_i(1)\alpha_i + \phi_i(L)z_{ip} \), \( \lambda_i = \phi_i(1)\beta_i \) and \( y_{it} = (y_{it-1}, ..., y_{it-p})' \) is the vector of lags with \( \Phi_i = (\phi_{1i}, ..., \phi_{pi})' \) being the associated vector of slope coefficients. If there is no trend,
\( \beta_i = 0 \) and so \( w_{it} = y_{it} - \Phi'_t y_{it} - \mu_i \). In any case, by using (1) to (3),
\[
\Delta w_{it} = c_i w_{i,t-1} + \epsilon_{it}
\]  
(6)
or, in terms of the observed variable,
\[
y_{it} = y_{it} - \Delta w_{it} + c_i w_{i,t-1} + \epsilon_{it} = y_{it-1} + \Phi'_t y_{it} + \lambda_i + c_i w_{i,t-1} + \epsilon_{it}.
\]
Thus, letting \( F_{i-1} \) denote the information set available at time \( t-1 \),
\[
E(y_{it}|F_{i-1}) = y_{it-1} + \Phi'_t y_{it} + \lambda_i + c_i w_{i,t-1}
\]
and
\[
\text{var}(y_{it}|F_{i-1}) = \omega_c^2 w_{i,t-1}^2 + \sigma_i^2,
\]
which can be used to obtain the log-likelihood function \( L \) of \( y_{i,p+1}, ..., y_{iT} \). In particular, suppose that \( \epsilon_{it} \) is normal, then, apart from constants,
\[
L = -\frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \ln \left( \text{var}(y_{it}|F_{i-1}) \right) - \frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{(y_{it} - E(y_{it}|F_{i-1}))^2}{\text{var}(y_{it}|F_{i-1})}
\]
\[
= -\frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \ln (\omega_c^2 w_{i,t-1}^2 + \sigma_i^2) - \frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{((\epsilon_i - \mu_c)w_{i,t-1} + \epsilon_{it})^2}{\omega_c^2 w_{i,t-1}^2 + \sigma_i^2}.
\]  
(7)
In Appendix A we show that under \( H_0 \) the log-likelihood is maximized by
\[
\sigma_i^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} (\Delta w_{it})^2,
\]
and that the Gradient and Hessian with respect to \( \mu_c \) and \( \omega_c^2 \) are given by
\[
g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left[ \frac{\Delta \hat{\epsilon}_{it}\hat{\epsilon}_{i,t-1}}{\frac{1}{3}((\Delta \hat{\epsilon}_{it})^2 - 1)\hat{\epsilon}_{i,t-1}^2} \right]
\]
and
\[
H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix} = -\sum_{i=1}^{N} \sum_{t=p+1}^{T} \left[ \frac{-\hat{\epsilon}_{i,t-1}^2}{\Delta \hat{\epsilon}_{i,t}^2} \frac{\hat{\epsilon}_{i,t-1}^2}{2} \frac{2((\Delta \hat{\epsilon}_{it})^2 - 1)\hat{\epsilon}_{i,t-1}^2}{\Delta \hat{\epsilon}_{i,t}^2} \right],
\]
respectively, where \( \hat{\epsilon}_{it} = w_{it}/\delta_i \). We also show that when properly normalized by \( N \) and \( T \) the Hessian is asymptotically diagonal. Thus, if all the parameters but \( \sigma_i^2 \) are known, then the Lagrange multiplier test statistic can be written as
\[
LM = g'(-H)^{-1} g = ALM + o_p(1),
\]
where
\[ ALM = -\frac{g_1}{H_{11}} - \frac{g_2}{H_{22}} = \left( \sum_{i=1}^{N} \sum_{t=p+1}^{p+2} \Delta \hat{e}_{it} \hat{e}_{it-1} \right)^2 + \frac{\left( \sum_{i=1}^{N} \sum_{t=p+2}^{T} \left( (\Delta \hat{e}_{it})^2 - 1 \right) \hat{e}_{it-1}^2 \right)^2}{2 \sum_{i=1}^{N} \sum_{t=p+2}^{T} (2(\Delta \hat{e}_{it})^2 - 1) (\hat{e}_{it-1})^4}, \]
which can be interpreted as an asymptotic Lagrange multiplier test statistic.

The formula for \( ALM \) is very simple and intuitive. In fact, a careful inspection reveals that the first part is nothing but the Lagrange multiplier test statistic for testing the null that \( \mu_c = 0 \) given \( \omega^2_c = 0 \). That is, the first part is the Lagrange multiplier unit root statistic based on the assumption of an homogenous \( \rho_i \). The second part is the Lagrange multiplier statistic for testing the null that \( \omega^2_c = 0 \) given \( \mu_c = 0 \).

The formula also reveals some interesting similarities with results obtained previously in the literature. In particular, note how the first part is the squared equivalent of the panel unit root test considered by Levin \textit{et al.} (2002).\(^3\) The second has no direct resemblance of anything that has been proposed earlier in the panel unit root literature. However, it can be seen as a panel version of the test statistic of Leybourne \textit{et al.} (1996), who consider the problem of testing the null of a fixed unit root against the randomized alternative in the context of a single time series. The test statistic as a whole can be regarded as a panel extension of the time series statistics discussed in Distasio (2008) and Ling (2004).

Even when \( \epsilon_{it} \) is normal the exact distribution of \( ALM \) is untractable. In this paper we therefore use asymptotic theory to obtain the limiting distribution of \( ALM \) as \( N, T \to \infty \).

The asymptotic null distribution of \( ALM \) is given in the following theorem.

\textbf{Theorem 1.} Under \( H_0 \) and Assumptions 1, 3 and 4,
\[ ALM \to_d X^2 + \frac{5}{24}(\kappa - 1) Y^2, \]
where \( X^2 \) and \( Y^2 \) are independent chi-squared random variables with one degree of freedom each.

\textbf{Remarks.}

(a) The theorem shows that \( ALM \) has the same limiting distribution in both models considered, and that this distribution is free of nuisance parameters, except for the dependence on \( \kappa \), the average fourth normalized moment of \( \epsilon_{it} \). If \( \epsilon_{it} \) is normal or if \( \kappa = 3 \),\(^3\) the first part of \( ALM \) can also be regarded as a panel version of the Lagrange multiplier unit root tests proposed in the time series literature by for example Ahn (1993) and Schmidt and Phillips (1992).
then \((\kappa - 1) = 2\) and hence the asymptotic distribution of \(ALM\) reduces to \(X^2 + \frac{5}{12} Y^2\). Thus, normality, or more generally, \(\kappa = 3\) implies a test distribution that is completely free of nuisance parameters.

(b) It is interesting to compare the asymptotic distribution of \(ALM\) with that obtained by Ling (2004) when testing for a unit root in a first-order autoregressive model with conditional heteroskedasticity, which can be reformulated as a random coefficient autoregressive model. The distribution of this test for cross-sectional unit \(i\) without any deterministic components is in our notation given by

\[
\frac{\left(\int_0^1 W_i(r)dW_i(r)\right)^2}{\int_0^1 W_i(r)^2dr} + \left(\kappa_i - 1\right)\frac{\left(\int_0^1 W_i(r)^2dV_i(r)\right)^2}{2\int_0^1 W_i(r)^4dr},
\]

where \(W_i(r)\) and \(V_i(r)\) are two independent standard Brownian motions on \(r \in [0, 1]\). The asymptotic distribution of our statistic can be regarded as

\[
\lim_{N \to \infty} \frac{\left(\frac{1}{N} \sum_{i=1}^N \int_0^1 W_i(r)dW_i(r)\right)^2}{\int_0^1 \sum_{i=1}^N W_i(r)^2dr} + \left(\kappa_i - 1\right)\frac{\left(\frac{1}{N} \sum_{i=1}^N \int_0^1 W_i(r)^2dV_i(r)\right)^2}{2\int_0^1 \sum_{i=1}^N W_i(r)^4dr}.
\]

Thus, by just comparing these two distributions, we see that the main effect of summing over the cross-sectional dimension is to smooth out the Brownian motion dependency for each unit.

(c) The requirement that \(\frac{N}{T} \to 0\) as \(N, T \to \infty\) is needed because while \(\Phi_i, \mu_i\) and \(\lambda_i\) are assumed to be known, \(\sigma_i^2\) is not and therefore has to be estimated.

Next we summarize the results obtained under \(H_1\).

**Theorem 2.** Under \(H_1\) and Assumptions 1, 3 and 4,

\[
ALM \to_d \frac{\mu_i^2}{2} + \mu_c \sqrt{2} X + X^2 + \frac{5}{24} (\kappa - 1) Y^2,
\]

where \(X\) and \(Y\) are as in Theorem 1.

**Remarks.**

(a) The first thing to note is that \(\omega_c^2\) does not enter the asymptotic distribution of the test. The reason for this originates with the rate of shrinking of the local alternative, which is determined by the normalization of the test statistic. With a composite test statistic like
ours, unless the normalization of the different parts is the same, the rate of shrinking of the local alternative is given by the lowest of the normalizing orders. In our case, the appropriate normalization for the first part of the test statistic is given by \( \frac{1}{\sqrt{NT}} \), while the normalization of the second part is \( \frac{1}{\sqrt{NT^{3/2}}} \). The rate of shrinking is therefore just enough to manifest \( \mu_c \) as a nuisance parameter in the asymptotic distribution of the first part of the statistic. The normalizing order of the second part, which represents the test of \( \omega_c^2 = 0 \), is higher and \( \omega_c^2 \) is therefore kicked out.

(b) The specification of \( H_1 \) has two effects. The first is to shift the mean of the limiting distribution of the test. In particular, since \( \mu_c^2 > 0 \), this means that the mean shifts to the left as we move away from \( H_0 \), suggesting that the test is unbiased and that its asymptotic local power therefore is greater than the size. The second effect, which is captured by \( \mu_c \sqrt{2} X \sim N(0, 2\mu_c^2) \), is to increase the variance of the limiting distribution. This effect is especially noteworthy as usually there is only the mean effect.

### 3.2 The feasible Lagrange multiplier test statistic

All results reported so far are based on the assumption that \( \Phi_i, \mu_i \) and \( \lambda_i \) are all known, which is of course not very realistic. Let us therefore consider using

\[
\hat{\omega}_{it} = y_{it} - \hat{\Phi}'_iy_{it} - \hat{\mu}_i - \hat{\lambda}_i(t - p)
\]  

as an estimator of \( \omega_{it} \), where \( \hat{\mu}_i = y_{ip+1} - \hat{\Phi}'_iy_{ip} - \hat{\lambda}_i \) with \( \hat{\lambda}_i \) and \( \hat{\Phi}_i \) being the least squares estimators of \( \lambda_i \) and \( \Phi_i \), respectively, in the first-differenced regression

\[
\Delta y_{it} = \lambda_i + \Phi_i' \Delta y_{it} + \epsilon_{it},
\]  

which is (5) with \( H_0 \) imposed.\(^4\) If there is no trend, then we remove the intercept, and compute \( \hat{\omega}_{it} = y_{it} - \hat{\Phi}'_iy_{it} - \hat{\mu}_i \), where \( \hat{\mu}_i = y_{ip+1} - \hat{\Phi}'_iy_{ip} \).\(^5\) The feasible Lagrange multiplier statistic in this model is given by

\[
FLM_1 = \frac{(\sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{\epsilon}_{it} \hat{\epsilon}_{it-1})^2}{\sum_{i=1}^{N} \sum_{t=p+2}^{T} \hat{\epsilon}_{it}^2} + \frac{12(\sum_{i=1}^{N} \sum_{t=p+2}^{T} ((\Delta \hat{\epsilon}_{it})^2 - 1) \hat{\epsilon}_{it-1}^2)^2}{5(k - 1) \sum_{i=1}^{N} \sum_{t=p+2}^{T} \hat{\epsilon}_{it}^2}.
\]

---

\(^4\) As shown in Lemma A.1 of Appendix A, under the null hypothesis \( \hat{\mu}_t, \lambda, \) and \( \hat{\Phi}_i \) are the feasible maximum likelihood estimators of \( \mu, \lambda, \) and \( \Phi_i \), respectively.

\(^5\) If in addition there is no serial correlation, then \( \hat{\omega}_{it} = y_{it} - \hat{\mu}_i \) with \( \hat{\mu}_i = y_{i1} \).
where $\hat{e}_{it} = \hat{w}_{it}/\hat{\sigma}_i$, $\hat{\sigma}_i^2 = 1/T - p - 1 \sum_{t=p+2}^T (\Delta \hat{w}_{it})^2$ and $\hat{k} = 1/(N(T - p - 1)) \sum_{i=1}^N \sum_{t=p+2}^T (\Delta \hat{w}_{it})^4 / \hat{\sigma}_i^4$. The reason for the subscript 1 is to indicate that the statistic has been computed for a particular choice of model, and that the limiting distribution depends on it. The asymptotic distribution of $FLM_1$ under $H_0$ is given in the following corollary.

**Corollary 1.** Under the conditions of Theorem 1,

$$FLM_1 \rightarrow_d X^2 + Y^2.$$  

Corollary 2 provides the corresponding result under $H_1$.

**Corollary 2.** Under the conditions of Theorem 2,

$$FLM_1 \rightarrow_d \mu_c^2 + \mu_c \sqrt{2} X + X^2 + Y^2.$$  

**Remarks.**

(a) The first term in the formula for $FLM_1$ is just the feasible version of the corresponding term in the formula for $ALM$ and does not require any explanation. The second term, however, is not as obvious. In Appendix B we show that as $N, T \rightarrow \infty$ with $N/T \rightarrow 0$

$$\frac{1}{NT^3} \sum_{i=1}^N \sum_{t=p+2}^T (2(\Delta \hat{e}_{it})^2 - 1)(\hat{e}_{it-1})^4 = \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=p+2}^T (\Delta \hat{e}_{it})^2 \hat{e}_{it-1}^4 + o_p(1) \rightarrow_p 1,$$

while $\frac{1}{\sqrt{NT^2}} \sum_{i=1}^N \sum_{t=p+2}^T ((\Delta \hat{e}_{it})^2 - 1)\hat{\sigma}_i^2 \rightarrow_d \sqrt{\frac{5}{12}(\kappa - 1)} Y$, which is the same limit as for the numerator of the second term in the formula for $ALM$. The second term in the formula for $FLM_1$ is therefore asymptotically equivalent to $\frac{24}{5}(\kappa - 1)$ times the corresponding term in $ALM$.

(b) As we point out in remark (a) above, $FLM_1$ is scale equivalent to $ALM$. This is very interesting because typically demeaning leads to an asymptotic bias that has to be removed in order to prevent the statistic from diverging, see for example Levin et al. (2002) and Im et al. (2003). We also see that the demeaning has no effect on the local power. This result is in agreement with the work of Moon et al. (2007), who develop a point optimal test statistic for the null that $\mu_c = 0$. According to their results estimation of intercepts does not affect maximal achievable power.6

6Unfortunately, the optimality property of the single parameter case does not translate directly to the present multiparameter case. The problem lies in that optimality for the single parameter case follows from maximizing power in the only direction available under the alternative hypothesis. In our case we have a power surface defined over all possible values of $\mu_c$ and $\omega_2^c$, and hence there is no obvious direction that should be used to maximize power.
It is interesting to compare the local power of the new test with the local power of the $Z_{\bar{t}}$ test of Im et al. (2003) and the $t^*_\delta$ test of Levin et al. (2002), two of its most natural competitors. As Moon and Perron (2008) show, under $H_1$ the latter statistic converges in distribution to $\frac{3}{2} \sqrt{\frac{5}{31}} \mu_c + N(0,1)$. The corresponding result for the former statistic is given in Harris et al. (2008) and is shown to be $0.282 \mu_c + N(0,1)$, where $\frac{3}{2} \sqrt{\frac{5}{31}} > 0.282$, suggesting that $t^*_\delta$ is most powerful. This can be seen in Figure 1, which plots the power of all three tests as a function of $\mu_c$.\(^7\) Intuitively, when one-directional alternatives are considered one-sided tests designed for that purpose should have the highest power. But when the alternative hypothesis moves in the direction of both $\mu_c \neq 0$ and $\omega^2_c > 0$, tests for the joint null hypothesis should have higher power. However, as the figure shows, except for the case when $-1.8 < \mu_c < 0$, $FLM_1$ is most powerful. The fact that the new test is most powerful even when the power is taken in the direction of only $\mu_c \neq 0$ is due to the rate of shrinking of the local alternative, which dominates the dependence upon $\omega^2_c$, thereby effectively making the test one-directional.

\(^7\)The figure is based on 5,000 replications.

Figure 1: Asymptotic local power as a function of $\mu_c$. 
Although unbiased in the case with a heterogeneous constant, the presence of a trend that needs to be estimated makes \( FLM_1 \) divergent. The source of this divergence is the numerator of the first term in the formula for \( FLM_1 \), which is no longer mean zero. In fact, as shown in Appendix C, \( \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{e}_{it} \hat{e}_{it-1} \to p - \frac{1}{2} \) as \( N, T \to \infty \) with \( \frac{N}{T} \to 0 \), suggesting that \( \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{e}_{it} \hat{e}_{it-1} \) diverges to negative infinity at rate \( \sqrt{N} \). But there is not only the mean effect, there is also a variance effect that works through the second term in the formula. Specifically, the estimation of the trend slope leads to an increase in variance, from \( \frac{1}{12} (\kappa - 1) \) in model 1 to \( \frac{1}{12} (\kappa - 1) \) in model 2.

However, this statistic has at least two drawbacks. Firstly, quite unexpectedly the usual practice of removing the nonzero mean of the statistic does not work in the sense that the asymptotic distribution of the mean-adjusted numerator of the first term of \( FLM_2 \) is degenerate. That is,

\[
\frac{1}{\sqrt{N T}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{e}_{it} \hat{e}_{it-1} + \frac{\sqrt{N}}{2} = o_p(1).
\]

In other words, the asymptotic null distribution of \( FLM_2 \) comes only from the second term in the formula. Secondly, and even more importantly, the test has no asymptotic power against \( H_1 \). Summarizing this, we have the following theorem.

**Theorem 3.** Under \( H_0 \) or \( H_1 \) and Assumptions 1, 3 and 4,

\[
FLM_2 \to_d X^2.
\]

**Remarks.**

(a) Since the asymptotic distribution under \( H_1 \) is the same as the one that applies under \( H_0 \), the local asymptotic power of \( FLM_2 \) is equal to the size. This stands in sharp contrast to the results obtained for \( ALM \) and \( FLM_1 \), which have nontrivial asymptotic power against \( H_1 \). This difference is a manifestation of the difficulty in detecting unit roots in the presence of heterogeneous trends, commonly referred to as the incidental trend problem, see Moon and Phillips (1999). The absence of local power is therefore not due to the degeneracy of the first term in the formula for \( FLM_2 \), which might otherwise seem like a very reasonable explanation.
The fact that ALM has nontrivial local power even in the presence of heterogeneous
trends suggests that the problem here is not the presence of trends per se but rather the
estimation thereof. Moon and Perron (2004, 2008), and Harris et al. (2008) consider the
effects of incidental trends when using least squares detrending. Theorem 3 extends
their results to the case of maximum likelihood demeaning.\footnote{Consistent with the results of Moon and Perron (2008), and Moon et al. (2006) our preliminary calculations suggest that, although absent under \( H_1 \), the new test has nontrivial power under alternatives that shrinks towards the null hypothesis at the slower rate of \( \sqrt{\frac{N}{T}} \).}

Despite the absence of local power, FLM\textsubscript{2} is consistent against a non-local alternative
in the sense that the probability of a rejection goes to one as \( N, T \to \infty \) for a set of
autoregressive parameters that does not depend on \( N \) or \( T \). The rate of the divergence
is \( \sqrt{NT} \), which is the same as for the Levin et al. (2002) and Im et al. (2003) tests.\footnote{A formal proof of this result can be obtained from the corresponding author.}

Although \( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{e}_{it} \hat{e}_{it-1} + \frac{\sqrt{N}}{2} \) is degenerate, \( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{e}_{it} \hat{e}_{it-1} + \frac{\sqrt{NT}}{2} \) is not. However, multiplication by \( \sqrt{T} \) introduces nuisance parameters that are
otherwise eliminated as \( T \to \infty \). It also makes the test dependent upon the distribution
of \( \epsilon_{it} \).

### 3.3 Generalizations

#### 3.3.1 Cross-section dependence

One drawback with the above analysis is that it supposes that the cross-sectional units are
independent, an assumption that is perhaps too strong to be held in many applications.
Accordingly, more recent panel unit root tests such as those of Bai and Ng (2004), Moon and
Perron (2004), Phillips and Sul (2003), and Pesaran (2007) relax this assumption by assuming
that the dependence can be represented by a common factor model. This approach fits very
well with the parametric flavor of our Lagrange multiplier framework, and it will therefore
be used also in this paper.

Suppose that \( \epsilon_{it} \) in (3) has the factor structure

\[
\epsilon_{it} = \Theta_i^\prime f_i + v_{it},
\]

where we assume for simplicity that \( f_i = (f_{1i}, ..., f_{ri})^\prime \) is an known \( r \)-dimensional vector of
common factors with \( \Theta_i = (\theta_{1i}, ..., \theta_{ri})^\prime \) being the associated vector of factor loadings, which
are assumed to be non-random. The error \( v_{it} \) is completely idiosyncratic. Both variables are assumed to satisfy Assumption 1 with \( f_t \) being independent of \( \Delta y_{it} \). Under these conditions, (6) becomes

\[
\Delta w_{it} = c_i w_{it-1} + \Theta'_i f_t + v_{it},
\]

which indicates that the feasible maximum likelihood estimator of \( \Theta_i \) in model 2 can be obtained by running the following least squares regression:

\[
\Delta y_{it} = \lambda_i + \Phi'_i \Delta y_{it} + \Theta'_i f_t + v_{it}. \tag{11}
\]

The factor-adjusted Lagrange multiplier test statistic is defined in exactly the same way as before but with \( \hat{w}_{it} \) given by

\[
\hat{w}_{it} = y_{it} - \hat{\Phi}_i' y_{it} - \hat{\mu}_i - \hat{\lambda}_i(t - p) - \hat{\Theta}_i' \sum_{s=p+2}^t f_s, \tag{12}
\]

where \( \hat{\mu}_i = y_{ip+1} - \hat{\Phi}_i' y_{ip} - \hat{\lambda}_i - \hat{\Theta}_i' f_{p+1} \) with \( \hat{\Phi}_i, \hat{\lambda}_i \) and \( \hat{\Theta}_i \) coming from the least squares fit of (11). The asymptotic distribution of this statistic is the same as the one given in Section 3.2 for the case with cross-section independence.

If \( f_t \) is also unknown, then we proceed as in Bai and Ng (2004), using the method of principal components to obtain consistent estimates. The trick is to note that under \( H_0 \), \( \Delta w_{it} = \Theta'_i f_t + v_{it} \), which is a nothing but a static common factor model in \( \Delta w_{it} \). In other words, had only \( \Delta w_{it} \) been known, we could have estimated \( f_t \) directly by the method of principal components. However, \( \Delta w_{it} \) is not known, and we must therefore apply the principal components method to \( \Delta \hat{w}_{it} \) instead, where \( \hat{w}_{it} \) is now as in (8). The testing can then be carried out as before but with \( f_t \) replaced by its principal components estimate.

Once this estimation process has been completed, there is of course no claim of validity of the resulting test, and we do not prove here that this approach is asymptotically valid. However, intuition suggests that it should perform well in practice, and unreported simulation evidence confirms this.

---

\(^{10}\) Of course, assuming that the common component enters via the serially uncorrelated error is by no means the only way in which the factor model can be specified. But it is convenient, see Pesaran (2007) for a detailed discussion of some alternative specifications.

\(^{11}\) There are two ways to eliminate the effects of the common component, depending on the estimation \( \Theta_i \). The first is the one described in the text, which amounts to replacing \( f_t \) by its principal components estimate, and then to estimate \( \Theta_i \) by applying least squares to the resulting first-difference regression. The second approach is to replace both \( f_t \) and \( \Theta_i \) by their principal components estimates. Unreported simulation evidence suggests that the first approach performs best.
3.3.2 Structural breaks

Analogous with the time series statistic studied by Amsler and Lee (1995), the asymptotic null distribution of $FLM^2$ computed under the assumption of a linear trend but no structural break is unaffected by the presence of a break in the level of $y_{it}$.

Let $D_t = 1(t > \tau)$, where $1(x)$ is the indicator function and $\tau$ indicates the timing of the break, which may be unit specific. The intuition behind the above result follows from the fact that $\frac{1}{\sqrt{T}} D_t = o_p(1)$, suggesting that the break has no effect on $\frac{1}{\sqrt{T}} \hat{\omega}_{it}$. Moreover, since $\Delta D_t = 0$ for all $t$ except when $t = \tau$, the effect on $\Delta \hat{\omega}_{it}$ is eliminated when subtracting the mean. The asymptotic null distribution of the $FLM^2$ is therefore unaffected.

The problem is that exclusion of the break makes the test biased towards accepting $H_0$. Thus, although the break does not affect the asymptotic null distribution of the test statistic, it does reduce its power. To avoid this, $D_t$ can be included in the analysis as an additional deterministic regressor, forming $d_{it} = \alpha_i + \beta_i (t - p) + \delta_i D_t$, where $\delta_i$ measures the magnitude of the break. The analysis can now be conducted exactly as before, augmenting (9) with $D_t$ as an additional regressor.

The development of procedures that accommodate breaks that are unknown is of interest but beyond the scope of the present contribution.

3.3.3 No restrictions on the relative expansion rate of $N$ and $T$

In applications when $N > T$ Assumption 4 no longer provides a reasonable approximation. In such cases we need to restrict the degree of heterogeneity that can be allowed. One way would be to assume that the heterogeneity can be regarded as random noise around an otherwise fixed mean value. But this induces a dependence on the distribution of the noise, which then has to be correctly specified. In this section we therefore go all the way and assume that $\alpha_i$, $\beta_i$, $\phi_i(L)$ and $\sigma_i^2$ are completely homogeneous across $i$.

The resulting test statistic is computed as before but with $\hat{\omega}_{it}$ given by

$$\hat{\omega}_{it} = y_{it} - \hat{\Phi}' y_{it} - \hat{\mu} - \hat{\beta} \left( (t - p) - \sum_{j=1}^{p} \hat{\phi}_j (t - p - j) \right),$$

(13)

where $\hat{\mu} = \bar{y}_{p+1} - \hat{\Phi}' \bar{y}_p - \hat{\lambda}$ with $\bar{y}_{p+1} = \frac{1}{N} \sum_{i=1}^{N} y_{ip+1}$ and an analogous definition of $\bar{y}_p$. Here $\hat{\lambda}$ and $\hat{\Phi} = (\hat{\phi}_1, ..., \hat{\phi}_p)'$ are the pooled least squares intercept and slope estimators in a regression of $\Delta y_{it}$ onto a constant and $\Delta y_{it}$. The homogenous trend slope $\beta$ cannot be
estimated directly, but it can be inferred from \( \hat{\lambda} \) and \( \hat{\Phi} \). Specifically, by using \( \lambda = \phi(1)\beta \) and a first-order Taylor expansion, it is not difficult to see that \( \hat{\beta} = \frac{\lambda}{(1-\sum_{j=1}^{r} \phi_j)} \) should be consistent for \( \beta \).

Replacing \( \hat{\sigma}^2 \) with \( \hat{\sigma}^2 = \frac{1}{N(T-p-1)} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\Delta \hat{w}_{it})^2 \), the feasible homogenous Lagrange multiplier test statistic in models 1 and 2 has the same form as \( FLM_1 \), and the asymptotic distributions under \( H_0 \) and \( H_1 \) are the same as the ones given in Corollaries 1 and 2, respectively. Thus, imposing homogeneity of the trend coefficient not only relaxes Assumption 4 but also removes the incidental trends problem and the absence of local power in \( \frac{1}{\sqrt{NT}} \) neighborhoods.

4 Simulations

In this section, we investigate the small-sample properties of the new test through a small simulation study using (1)–(3) to generate the data. For simplicity, we assume that \( \phi_i(L) = 1 - \phi L, \alpha_i = \beta_i = 1 \) and \( \varepsilon_{it} \sim N(0,1) \).

A total of seven configurations of the autoregressive parameter \( \rho_i \) and the drift parameter \( c_i \) are considered, where the latter is assumed to be generated as \( c_i \sim U(a, b) \). The first configuration is for analyzing the size of the test, while the remaining six are for analyzing the power. Three of these are local to the null hypothesis and three are non-local. Specifically, the following cases are considered:

1. \( \rho_i = 1 \) for all \( i \);
2. \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i = -10 \) for all \( i \);
3. \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i \sim U(-20,0) \);
4. \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i \sim U(-40,20) \);
5. \( \rho_i = 1 + c_i \) with \( c_i = -0.05 \) for all \( i \);
6. \( \rho_i = 1 + c_i \) with \( c_i \sim U(-0.1,0) \);
7. \( \rho_i = 1 + c_i \) with \( c_i \sim U(-0.15,0.05) \).

Note that with this specification of \( c_i \), \( \mu_c = \frac{1}{2}(a + b) \) and \( \omega_c^2 = \frac{1}{12}(a - b)^2 \). Hence, \( \mu_c \) is the same in cases 2–4, and also in cases 5–7. The only thing that separates for example case
1 from cases 2 and 3 is therefore the variance, which goes from zero in case 1 to 33.33 in case 2 to 300 in case 3. This direction away from the null is interesting to consider since in our random coefficient setting there is not just the mean of $c_i$ that matters but also the variance. The data in all six cases are generated for 5,000 panels with $T + 100$ time series observations, where the first 100 are disregarded to reduce the effect of the initial value, which is set to zero.

For the sake of comparison, the Levin *et al.* (2002) $t^*_\delta$ statistic and Im *et al.* (2003) $Z_{\bar{t}}$ statistic are also simulated. As explained earlier, both are constructed as $t$-ratios of the null that $\rho_i = 1$ for all $i$. The difference is that while $t^*_\delta$ is based on the $t$-ratio of the pooled least squares estimator of $\rho$, $Z_{\bar{t}}$ is based on the average of the individual $t$-ratios of the least squares estimator of $\rho_i$. As with the new test, $Z_{\bar{t}}$ is fully parametric with respect to the serial correlation properties of the data, and hence only requires lag augmentation. By contrast, $t^*_\delta$ not only requires lag augmentation but also semiparametric estimation of the so-called long-run variance of $u_{it}$.

In the simulations the lag length is selected using the Schwarz Bayesian information criterion, which facilitates a data dependent choice. Consistent with the results of Ng and Perron (1995), the maximum number of lags is allowed to increase with $T$ at the rate $4(T/100)^{2/9}$. To also allow for the possibility of heterogeneous lag lengths, the criterion is evaluated once for each unit. As for the semiparametric estimation needed for computing $t^*_\delta$, we follow the recommendation of Levin *et al.* (2002) and use the Bartlett kernel with the bandwidth parameter set equal to $3.21T^{1/3}$.

The $t^*_\delta$ and $Z_{\bar{t}}$ statistics can be constructed in two ways depending on the choice of mean and variance adjustment, which can be either asymptotic or sample-specific. Our test is asymptotic, suggesting that the most appropriate comparison here is obtained by using the former adjustments. However, since Im *et al.* (2003) do not provide any asymptotic results for their test, $t^*_\delta$ and $Z_{\bar{t}}$ are simulated based on the small-sample moments.\(^\text{12}\) For brevity, we only report the size and power at the 5% significance level. Also, since size accuracy is not perfect, all powers are adjusted so that each test has the same level of 5% when the null hypothesis is true. All computational work has been performed in GAUSS.\(^\text{13}\)

\(^{12}\)The use of sample-specific adjustment terms is expected to lead to better performance in the simulations, which is also what we find. The comparison with the Lagrange multiplier test is therefore biased in favor of its competitors.

\(^{13}\)In addition to the results reported here, we have experimented with a large number of different parame-
Consider first the size results for model 1, which are reported in Table 1. It is seen that among the three tests considered the best size accuracy is generally obtained by using $Z_{\bar{t}bar}$, with $FLM_1$ performing only marginally worse. In fact, our results suggest that these tests are remarkably robust even to quite high degrees of serial correlation, a valuable property that is not very common. Of course, the accuracy is not perfect, and some distortions remain. In particular, we see that there is a slight tendency for the test to become oversized as $N$ increases, although the distortions vanish quickly as $T$ increases. The $t_\delta^*$ test performs worst with massive size distortions, even when $\phi = 0$ and there is no serial correlation.

The results from the local power of the tests in cases 2–4 are even more encouraging. Indeed, as Table 1 shows, the new test is almost uniformly more powerful than the other tests, and this holds even when the power is taken in the direction of only $\mu_c < 0$, which is consistent with our asymptotic results on this point, as summarized by Figure 1. We also see that increasing the variance does not lead to any increase in power, as should be expected from remark (c) following Corollary 2. Moreover, although there is a small increase among the smaller values of $N$ and $T$, the power is quite flat in the sample size, which is in accordance with our expectations, since asymptotically there is no dependence on $N$ and $T$.

The poor performance of the $t_\delta^*$ test is due to overfitting, which apparently can cause drastic reductions in power. This is in agreement with the results of Westerlund (2009), who shows that the power of $t_\delta^*$ depends heavily on the choice of lag length, and to an even greater extent on the choice of bandwidth. The fact that the new test seems much more robust in this regard is of some importance from an applied standpoint because these are difficult choices.

Focusing now on the non-local power, we see that while $FLM_1$ keeps its relative advantage in cases 5 and 7, in case 6 $Z_{\bar{t}bar}$ is ranked first, although not by much. We also see that while the level of the power is higher than in cases 2–4, in relative terms the $t_\delta^*$ test is still dominated by the others. As indicated in remark (c) of Theorem 3, under the non-local alternatives considered here, all three tests diverge at rate $\sqrt{NT}$. In agreement with this we see that increasing values of $N$ and $T$ lead to roughly the same increase in power, and that the magnitude of the increase is roughly of the expected rate.

Consider next the results reported in Table 2 for model 2. The first thing to notice is the parameterizations of the data generating process, including negative autoregressive errors, moving average errors, and heterogeneous deterministic intercept and trend terms. Except possibly for the usual distortions in the case with negative moving average errors, the conclusions were not altered. These results are available from the corresponding author upon request. Some results based on alternative lag selection rules are also available.
power in cases 2–4, which is almost absent. The theoretical result that the distribution of the new statistic is the same under the null and local alternative hypotheses implies that the power should be roughly equal to size, or 5%. Our results are quite suggestive of this. The results for the non-local alternatives of cases 5–7 are more promising, but the power is still very low, especially among the smaller values of $N$ and $T$.

5 An empirical illustration

In a well-functioning market, an increase in demand brought about by for example higher income should be accompanied by a one-for-one increase in supply, with prices being left unchanged. By contrast, in a poorly functioning market, demand and supply do not move one-for-one and therefore prices rise. This is presently the situation in many housing markets around the world. In the United States, prices have grown so fast that it has raised fears of speculative bubbles with real prices moving away from real income.\footnote{These concerns culminated with the eruption of the sub-prime mortgage crisis in mid-2006, which have lead to plunging property prices and a slowdown in the United States economy.}

This development is illustrated in Figure 2, which plots the cross-sectional mean, range and normal 95% confidence bands for the log of the price-to-income ratio for 49 states between 1975 and 2003.\footnote{The data are taken from Holly et al. (2009), and include for each state the real house price and income, which are both transformed by taking logs.} As can be seen, the ratio first increased but then in the early 1980’s, a period largely consistent with the NBER business cycle peak of January 1980, it started to decline, levelling off in the early 1990’s. The sharp increase in the end of the sample is consistent with the NBER peak of March 2001.

Figure 2 suggests that if the absence of speculative bubbles is to be interpreted as a mean-reversion of the price-to-income ratio, then there is little evidence to support it. It is observations like this that have recently led many researchers to question the health of the United States housing market. One such study is that of Holly et al. (2009), in which the authors deduce evidence of a stable long-run one-to-one relationship between prices and income, suggesting that the market is actually in good health. However, their unit root test is based on the assumption that the data are integrated of at most order one, which naturally raises the question of how the conclusions hold up in case of a violation.\footnote{Preliminary evidence at the state level indicates that the fully stationary alternative is inappropriate. Take for example prices, for which the estimated first-order autoregressive coefficient can be as high as 1.17 for some states, suggesting the presence of explosive behavior.}
In this section, we try to shed some light on this issue by reevaluating the results of Holly et al. (2009) based on the new test. The appropriate number of lags to use is determined as in Section 4, using Schwarz Bayesian information criterion. Because of the strong co-movement in the data the test is implemented while allowing for up to four common factors with the exact number determined using the $IC_2$ criterion of Bai and Ng (2002).\footnote{In addition to using the principal components approach of Bai and Ng (2004) to estimate the factors we tried the cross-sectional average approach of Pesaran (2007) with very little differences in the results.} Most of the price and income series also seem to be trending, implying that model 1 with only an intercept might not provide an accurate description of the data. The approach taken here is very simple and is based on using the Ayat and Burridge (2000) approach to determine the significance of the individual trend slopes. Only if the zero slope hypothesis is accepted for all states do we conclude that model 1 is appropriate. The results suggest that for a majority of the states the zero slope hypothesis must be rejected. We therefore focus on model 2, but include the results for model 1 for comparison.

The results are presented in Table 3. In agreement with the findings of Holly et al. (2009), we see that the factor-adjusted test is unable to reject the unit root null at the 5% level for

Figure 2: The cross-sectional mean of the price-to-income ratio.
prices and income in their levels, but not in their first differences. The fact that the unad-
justed test always rejects is not totally unexpected given the well-known size distortions of
so-called first-generation panel unit root tests in the presence of unattended cross-section de-
pendence. The results for the price-to-income ratio, which also agree with Holly et al. (2009),
show that the variable is trend-stationary, suggesting the presence of a stable long-run one-
to-one relationship between prices and income.

6 Conclusion

This paper has developed a new procedure for testing the null hypothesis of a unit root
in panels where the heterogeneity of the autoregressive coefficient can be assumed to be
random across the cross-section. This is quite important since in most, if not all, related
work, whenever heterogeneity is allowed, it is assumed to be non-random. This means that
each individual coefficient has to be estimated separately, leading to excess variation in the
test. The purpose of the current paper was to device a test that exploits the information that
under the null hypothesis of a unit root, when a random approach is used, the autoregressive
coefficients have unit mean and zero variance. This led us naturally to the consideration of
the Lagrange multiplier, or score, principle.

We have shown that with individual constants, the proposed Lagrange multiplier test
has power in a local neighborhood that shrinks towards the null hypothesis at rate \( \frac{1}{\sqrt{NT}} \). The
limiting distribution of the new test statistic is chi-squared and therefore no special table is
required to compute \( p \)-values. We have also shown that in the presence of heterogeneous
trends that have to be estimated, although still consistent against non-local alternatives, the
local power of the test is equal to the size.

Finally, we have provided simulation evidence that supports our theoretical results. In
particular, we have shown that when no estimation of deterministic trends is necessary the
new test has good size accuracy and excellent power in comparison to other tests. When
such estimation is necessary, the test typically has no power beyond size.
References


Appendix A: Derivation of the true Lagrange multiplier statistic

In this appendix we derive the true Lagrange multiplier statistic. For brevity, the results are only provided for the case with a trend, which is the most general deterministic specification considered.

Lemma A.1. Under $H_0$ and Assumptions 1–3 in the model with a trend the maximum likelihood estimators of $\sigma_i^2$, $\mu_i$, $\lambda_i$ and $\Phi_i$ are given by

(a) $\hat{\sigma}_i^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \epsilon_{it}^2$, 
(b) $\hat{\mu}_i = y_{ip+1} - \Phi'_i y_{ip} - \lambda_i$, 
(c) $\hat{\lambda}_i = \Delta y_{it} - \Phi'_i \Delta y_{it}$, 
(d) $\hat{\Phi}_i = \left( \sum_{t=p+2}^{T} (\Delta y_{it} - \Delta \bar{y}_i)(\Delta y_{it} - \Delta \bar{y}_i)' \right)^{-1} \sum_{t=p+2}^{T} (\Delta y_{it} - \Delta \bar{y}_i)(\Delta y_{it} - \Delta \bar{y}_i)$,

where $\Delta y_i = \frac{1}{T-p-1} \sum_{t=p+1}^{T} \Delta y_{it}$ with an analogous definition of $\Delta \bar{y}_i$.

Proof of Lemma A.1.

Consider (a). With the trend specification of $d_{it}$,

$\phi_i(L)y_{ip+1} = \phi_i(1)(\alpha_i + \beta_i) + \rho_i \phi_i(L)z_{ip} + \epsilon_{ip+1}$

for $t = p + 1$, and for $t = p + 2, ..., T$,

$\phi_i(L)y_{it} = (1 - \rho_i) \phi_i(L)(\alpha_i + \beta_i(t-p)) + \rho_i \phi_i(L)(y_{it-1} + \beta_i) + \epsilon_{it}$.

Under $H_0$ these two equations reduce to

$\phi_i(L)y_{ip+1} = \phi_i(1)(\alpha_i + \beta_i) + \epsilon_{ip+1} = \mu_i + \lambda_i + \epsilon_{ip+1}, \quad \text{(A1)}$

$\phi_i(L)y_{it} = \phi_i(1)(y_{it-1} + \beta_i) + \epsilon_{it} = \phi_i(L)y_{it-1} + \lambda_i + \epsilon_{it}. \quad \text{(A2)}$

Moreover, under $H_0$ the log-likelihood function in (7) reduces to

$L = - \frac{T-p}{2} \sum_{i=1}^{N} \ln(\sigma_i^2) - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \sum_{t=p+1}^{T} \epsilon_{it}^2. \quad \text{(A3)}$

Clearly,

$$\frac{\partial L}{\partial \sigma_i^2} = - \frac{T-p}{2\sigma_i^2} + \frac{1}{2\sigma_i^2} \sum_{t=p+1}^{T} \epsilon_{it}^2,$$
which can be put equal to zero, and then solved for $\sigma_t^2$, proving (a).

Consider (b). By imposing $H_0$ and then concentrating with respect to $\sigma_t^2$,

\[
L = -\frac{T-p}{2} \sum_{i=1}^{N} \ln(\tilde{\sigma}_t^2) = -\frac{T-p}{2} \sum_{i=1}^{N} \ln\left(\frac{1}{T-p} \sum_{t=p+1}^{T} \tilde{\sigma}_t^2\right),
\]

(A4)

where, making use of (A1) and (A2),

\[
\sum_{t=p+1}^{T} \tilde{\sigma}_t^2 = \sum_{t=p+1}^{T} e_{ip+1}^2 + \sum_{t=p+2}^{T} e_{it}^2 = (\phi_t(L)y_{ip+1} - \mu_i - \lambda_i)^2 + \sum_{t=p+2}^{T} (\phi_t(L)\Delta y_{it} - \lambda_i)^2.
\]

It follows that

\[
\frac{\partial L}{\partial \mu_i} = \frac{1}{\tilde{\sigma}_t^2} \sum_{t=p+1}^{T} (\phi_t(L)\Delta y_{it} - \lambda_i),
\]

implying $\tilde{\mu}_i = \phi_t(L)y_{ip+1} - \lambda_i = y_{ip+1} - \Phi_t^{'}y_t - \lambda_i$.

Moreover, since $e_{ip+1}$, and therefore also $e_{ip+1}^2$, is zero when evaluated at $\mu_i = \tilde{\mu}_i$ and $\sigma_t^2 = \tilde{\sigma}_t^2$,

\[
\frac{\partial L}{\partial \lambda_i} = \frac{1}{\tilde{\sigma}_t^2} \sum_{t=p+1}^{T} (\phi_t(L)\Delta y_{it} - \lambda_i),
\]

giving $\tilde{\lambda}_i = \frac{1}{T-p} \sum_{t=p+1}^{T} \phi_t(L)\Delta y_{it} = \Delta \overline{y}_i - \Phi_t^{'}\Delta \overline{y}_t$, which establishes (c).

Similarly, by using $\tilde{\lambda}_i$ in place of $\lambda_i$,

\[
\sum_{t=p+1}^{T} \tilde{\sigma}_t^2 = \sum_{t=p+1}^{T} (\phi_t(L)\Delta y_{it} - \tilde{\lambda}_i)^2 = \sum_{t=p+2}^{T} ((\Delta y_{it} - \Delta \overline{y}_i) - \Phi_t^{'}(\Delta y_{it} - \Delta \overline{y}_i)) (\Delta y_{it} - \Delta \overline{y}_i)'^2,
\]

and so

\[
\frac{\partial L}{\partial \Phi_t} = \frac{1}{\tilde{\sigma}_t^2} \sum_{t=p+2}^{T} ((\Delta y_{it} - \Delta \overline{y}_i) - \Phi_t^{'}(\Delta y_{it} - \Delta \overline{y}_i)) (\Delta y_{it} - \Delta \overline{y}_i)',
\]

from which we deduce that

\[
\Phi_t = \left( \sum_{t=p+2}^{T} ((\Delta y_{it} - \Delta \overline{y}_i) (\Delta y_{it} - \Delta \overline{y}_i)')^{-1} \sum_{t=p+2}^{T} (\Delta y_{it} - \Delta \overline{y}_i) (\Delta y_{it} - \Delta \overline{y}_i)' \right)^{-1}
\]

This establishes (d) and hence the proof of the lemma is complete. 

\(\blacksquare\)

**Lemma A.2.** Under the conditions of Lemma A.1,

(a) \[ \frac{\partial L}{\partial \mu_i} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta \tilde{e}_{it} \tilde{e}_{i,t-1}. \]
\[
(b) \quad \frac{\partial^2 L}{(\partial \mu_c)^2} = -\sum_{i=1}^{N} \sum_{t=p+1}^{T} \tilde{e}_{it-1}^2 \\
(c) \quad \frac{\partial L}{\partial \omega_c^2} = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left( (\Delta \tilde{e}_{it})^2 - 1 \right) \tilde{e}_{it-1}^2 \\
(d) \quad \frac{\partial^2 L}{(\partial \omega_c^2)^2} = -\frac{1}{2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (2(\Delta \tilde{e}_{it})^2 - 1) \tilde{e}_{it-1}^4 \\
(e) \quad \frac{\partial^2 L}{\partial \mu_c \partial \omega_c^2} = -\sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta \tilde{e}_{it} \tilde{e}_{it-1}^3.
\]

**Proof of Lemma A.2.**

We prove (a). The proofs of (b) to (e) follow by similar arguments.

From (7),

\[
\frac{\partial L}{\partial \mu_c} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{(c_i - \mu_c)w_{it-1} + \epsilon_{it})w_{it-1}}{\omega_c^2 w_{it-1}^2 + \sigma_i^2}.
\]

By dividing both the numerator and the denominator by \(\sigma_i^2\), and then imposing \(H_0\), we obtain

\[
\frac{\partial L}{\partial \mu_c} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \frac{(c_i - \mu_c)\epsilon_{it-1} + \epsilon_{it} \epsilon_{it-1}}{\omega_c^2 \epsilon_{it-1}^2 + 1} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \epsilon_{it} \epsilon_{it-1} = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta \epsilon_{it} \epsilon_{it-1},
\]

where \(\epsilon_{it} = w_{it} / \sigma_i\) and \(\epsilon_{it} = \epsilon_{it} / \sigma_i\). The required result is obtained by concentrating the above expression with respect to \(\sigma_i^2\). \[\blacksquare\]

The Lagrange multiplier statistic is defined as

\[
LM = g' (-H)^{-1} g,
\]

(A5)

where by Lemma A.2,

\[
g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial \mu_c} \\ \frac{\partial L}{\partial \omega_c} \end{bmatrix}, \quad H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 L}{\partial \mu_c^2} & \frac{\partial^2 L}{\partial \mu_c \partial \omega_c} \\ \frac{\partial^2 L}{\partial \mu_c \partial \omega_c} & \frac{\partial^2 L}{\partial (\omega_c)^2} \end{bmatrix}.
\]

We now show that when properly normalized \(H\) is asymptotically diagonal, which yields the desired result after substituting for \(g\) and \(H\) in (A5). Let us therefore consider

\[
LM = -(G^{-1} g)' (G^{-1} H G^{-1})^{-1} (G^{-1} g),
\]

28
where

\[ G = \begin{bmatrix} \sqrt{NT} & 0 \\ 0 & \sqrt{NT^3/2} \end{bmatrix} \]

is the normalizing matrix. The off-diagonal element of \( G^{-1} H G^{-1} \) is given by

\[
\frac{1}{NT^5/2} H_{12} = -\frac{1}{NT^5/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta \epsilon_i \epsilon_{it-1} = -\frac{1}{NT^5/2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 \epsilon_i \epsilon_{it-1}
\]

\[
= \frac{1}{NT^5/2} H_{12}^0 + R,
\]

where \( H_{12}^0 = \sum_{i=1}^{N} \sum_{t=p+1}^{T} \epsilon_i \epsilon_{it}^3 \), \( R = \frac{1}{NT^5/2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (w_i^2 - 1) \epsilon_i \epsilon_{it}^3 \), \( w_i = \sigma_i^2 / \delta_i^2 \) and \( s_{it} = \sum_{k=p+1}^{t} \epsilon_{ik} \).

Consider \( H_{12}^0 \). Since \( \epsilon_{it} \) is independent of \( s_{it-1} \) as well as across both \( i \) and \( t \),

\[
\frac{1}{NT^5/2} E(H_{12}^0) = \frac{1}{NT^5/2} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\epsilon_{it}) E(s_{it-1}) = 0.
\]

Also, by a functional central limit theorem, \( \frac{1}{\sqrt{T}} s_{it-1} = O_p(1) \), which in turn suggests that a central limit theorem should apply to \( \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \epsilon_i \epsilon_{it}^3 \). Hence,

\[
\frac{1}{NT^5/2} H_{12} = O_p \left( \frac{1}{\sqrt{NT}} \right)
\]

and by the Cauchy–Schwarz inequality,

\[
R = \frac{1}{NT^5/2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (w_i^2 - 1) \epsilon_i \epsilon_{it}^3 \leq \left[ \frac{1}{N} \sum_{i=1}^{N} (w_i^2 - 1)^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^5/2} \sum_{t=p+1}^{T} \epsilon_i \epsilon_{it}^3 \right)^2 \right]^{1/2} = O_p \left( \frac{1}{T} \right),
\]

where we have made used of the fact that \( w_i - 1 = O_p(1/\sqrt{T}) \), as follows from a first-order Taylor expansion of the inverse of \( \delta_i^2 \), which is such that

\[
\delta_i^2 = \frac{1}{T-\rho} \sum_{t=p+1}^{T} \epsilon_{it}^2 = \sigma_i^2 + O_p \left( \frac{1}{\sqrt{T}} \right). \tag{A6}
\]

In fact, we even have \( E(\delta_i^2) = \frac{1}{T-\rho} \sum_{t=p+1}^{T} E(\epsilon_{it}^2) = \sigma_i^2 \), meaning that \( \delta_i^2 \) is not only consistent but also unbiased.

It follows that

\[
\frac{1}{NT^5/2} H_{12} = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{T} \right),
\]

29
proving that the Hessian is indeed asymptotically diagonal. Lemma A.3 further shows that
minus the Hessian for \( \sigma_i^2, \mu_i, \lambda_i \) and \( \Phi_i \) tends to a positive definite matrix, verifying that \( \sigma_i^2, \mu_i, \lambda_i \) and \( \Phi_i \) maximizes the log-likelihood function.

**Lemma A.3.** Under the conditions of Lemma A.1, as \( T \to \infty \)

\[-(G^*)^{-1}H(G^*)^{-1} \to H^0 > 0,\]

where

\[
H^0 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & \sigma_i^2 \text{cov}(\Delta y_{it})
\end{bmatrix}, \quad G^* = \begin{bmatrix}
\sqrt{T} & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & \sqrt{T} & 0 \\
0 & 0 & 0 & \sqrt{T}
\end{bmatrix},
\]

**Proof of Lemma A.3.**

From the proof of Lemma A.1 we have that when evaluated at \( \sigma_i^2, \mu_i, \lambda_i \) and \( \Phi_i \) minus the Hessian becomes

\[
-H^* = \frac{1}{\sigma_i^4} \begin{bmatrix}
\frac{1}{2}(T - p) & 0 & 0 & 0 \\
0 & 1 & 1 & \sigma_i^2 y_{ip} \\
0 & 1 & T - p - 1 & \sigma_i^2(T - p - 1) \Delta y_i \\
0 & \sigma_i^2 y_{ip} & \sigma_i^2(T - p - 1) \Delta y_i & \sigma_i^2 \sum_{t=p+2}^T (\Delta y_{it} - \Delta y_i) (\Delta y_{it} - \Delta y_i)'
\end{bmatrix}
\]

\[
= \frac{1}{\sigma_i^4} G^* D G^*,
\]

where

\[
D = \begin{bmatrix}
\frac{1}{2}(1 - \frac{p}{T}) & 0 & 0 & 0 \\
0 & 1 & \frac{1}{\sqrt{T}} & \frac{1}{\sqrt{T}} \sigma_i^2 y_{ip}' \\
0 & \frac{1}{\sqrt{T}} & 1 - \frac{p+1}{T} & \sigma_i^2(1 - \frac{p+1}{T}) \Delta y_i' \\
0 & \frac{1}{\sqrt{T}} \sigma_i^2 y_{ip}' & \sigma_i^2(1 - \frac{p+1}{T}) \Delta y_i' & \sigma_i^2 \sum_{t=p+2}^T (\Delta y_{it} - \Delta y_i) (\Delta y_{it} - \Delta y_i)'
\end{bmatrix}.
\]

Note that \( G^* > 0 \). Thus, by using the results of Abadir and Magnus (2005), if we can show that \( D > 0 \), then \(-H^* > 0\). Towards this end, note that \( D \to H^0 \) as \( T \to \infty \), where \( \text{cov}(\Delta y_{it}) \) is a diagonal matrix, which implies \( \det(\text{cov}(\Delta y_{it})) > 0 \). Hence, because all the leading principal minors of \( H^0 \) are positive definite, \( H^0 > 0 \).
Appendix B: Asymptotic properties of the true Lagrange multiplier statistic

Proof of Theorem 1.

Write

\[ ALM = \left( \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta \varepsilon_{it} \varepsilon_{it-1} \right)^2 + \left( \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\Delta \varepsilon_{it})^2 - 1 \right) \varepsilon_{it-1}^2 \]

\[ = \left( \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i \varepsilon_{it} s_{it-1} \right)^2 + \left( \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 (\varepsilon_{it}^2 - \varepsilon_{it-1}^2) s_{it-1}^2 \right) \]

where \( w_i \) and \( s_{it} \) are as in Appendix A.

Consider \( I \), which we write as

\[ I = \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i \varepsilon_{it} s_{it-1} \right)^2 = I_1^2 + I_2, \]

where \( I_1 = I_1^0 + R_1 \) with \( I_1^0 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \varepsilon_{it} s_{it-1} \)

\[ R_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (w_i - 1) \varepsilon_{it} s_{it-1} \]

\[ \leq \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (w_i - 1) \right]^{1/2} \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=p+1}^{T} \varepsilon_{it} s_{it-1} \right) \right]^{1/2} = O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right), \]

which goes to zero if \( \frac{N}{T} \to 0 \) as \( N, T \to \infty \). It follows that \( I_1 \) is asymptotically equivalent to \( I_1^0 \), whose expectation is given by

\[ E(I_1^0) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} E(\varepsilon_{it}) E(s_{it-1}) = 0. \]

The computation of the variance is simplified by noting that as \( T \to \infty\)

\[ \frac{1}{T} \sum_{t=p+1}^{T} \varepsilon_{it} s_{it-1} \to_d \int_{0}^{1} W_i(r) dW_i(r), \]

where \( W_i(r) \) is a standard Brownian motion on \( r \in [0, 1] \). Thus, by the continuous mapping
which, via the continuous mapping theorem, gives 
\[
\text{var}(I^1_i) \to E \left[ \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \int_0^T W_i(r) dW_i(r) \right)^2 \right]
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \int_0^T \int_0^T E(W_i(r)W_j(u))E(dW_i(r)dW_j(u))
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \int_0^T E(W_i(r)^2)E(dW_i(r)^2) = \int_0^T rdr = \frac{1}{2},
\]
where the second equality follows from the fact that \(E(dW_i(r)dW_j(u)) = 0\) for all \(i \neq j\) and \(r \neq u\), while the third uses \(E(W_i(r)^2) = r\) and \(dW_i(r)^2 = dr\).

Define \(X_i = \frac{\sqrt{2}}{\sqrt{NT}} \sum_{t=p+1}^{T} \varepsilon_t s_{it-1}\), which is independent across \(i\) with mean zero and variance \(\text{var}(X_i) = O(1/N)\). Therefore, according to Theorem 2 of Phillips and Moon (1999), if we can show that for all \(\delta > 0\),
\[
\lim_{N,T \to \infty} \sum_{i=1}^{N} E(X_i^2 1(|X_i| > \delta)) = 0,
\]
where \(1(x)\) is the indicator function, then \(\sum_{i=1}^{N} X_i \to_d X \sim N(0,1)\) as \(N, T \to \infty\).

To verify this condition we make use of the Cauchy–Schwarz inequality, which yields
\[
E(X_i^2 1(|X_i| > \delta)) \leq \sqrt{E(X_i^4)E(1(|X_i| > \delta))}
\]
and by further application of the Markov inequality, \(E(1(|X_i| > \delta)) \leq \frac{1}{\delta} E(X_i^2)\). Thus,
\[
\sum_{i=1}^{N} E(X_i^2 1(|X_i| > \delta)) \leq \frac{1}{\delta} \sum_{i=1}^{N} \sqrt{E(X_i^4)E(X_i^2)} \leq \frac{1}{\delta} \left( \sum_{i=1}^{N} E(X_i^4) \right)^{1/2} \left( \sum_{i=1}^{N} E(X_i^2) \right)^{1/2}
\]
\[
= O \left( \frac{1}{\sqrt{N}} \right).
\]
Therefore, since the above condition holds, as \(N, T \to \infty\)
\[
I^0_1 = \frac{1}{\sqrt{2}} \sum_{i=1}^{N} X_i \to_d \frac{1}{\sqrt{2}} X,
\]
which, via the continuous mapping theorem, gives \((I^0_1)^2 \to_d \frac{1}{2} X^2\).

Consider \(I^0_2 = \frac{1}{NT^2} \sum_{t=1}^{N} \sum_{i=p+1}^{T} s_{it-1}^2\). As \(T \to \infty\)
\[
\frac{1}{T^2} \sum_{t=p+1}^{T} s_{it-1}^2 \to_w \int_0^1 W_i(r)^2 dr,
\]

32
and therefore

\[ E(I_2) \rightarrow \frac{1}{N} \sum_{i=1}^{N} \int_0^1 E(W_i(r)^2)dr = \int_0^1 r dr = \frac{1}{2}. \]

Thus, by Corollary 1 of Phillips and Moon (1999), if \( \frac{1}{N} \sum_{t=p+1}^{T} s_{it-1}^2 \) is uniformly integrable in \( T \), then \( I_2 \rightarrow_p \frac{1}{2} \) as \( N, T \rightarrow \infty \). But \( \frac{1}{N} \sum_{t=p+1}^{T} s_{it-1}^2 \rightarrow_w \int_0^1 W_i(r)^2 dr \), and therefore uniform integrability is a direct consequence of

\[ E \left( \frac{1}{T^2} \sum_{t=p+1}^{T} s_{it-1}^2 \right) = tr \left( \frac{1}{T^2} \sum_{t=p+1}^{T} E(s_{it-1}^2) \right) \rightarrow E \left( \int_0^1 W_i(r)^2 dr \right), \]

see Appendix C of Phillips and Moon (1999).

Hence, because \( I_2 = I_2^p + R_2 \), where

\[ R_2 = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (w_i - 1)s_{it-1}^2 \]

\[ \leq \left[ \frac{1}{N} \sum_{i=1}^{N} (w_i - 1)^2 \right]^{1/2} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T^2} \sum_{t=p+1}^{T} s_{it-1}^2 \right)^2 \right]^{1/2} = O_p \left( \frac{1}{\sqrt{T}} \right), \]

by Taylor expansion and then passing \( N, T \rightarrow \infty \) with \( \frac{N}{T} \rightarrow 0 \), we obtain

\[ I = I_2^p = \frac{(I_2^p)^2}{I_2} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) \rightarrow_d X^2. \]  

(A8)

Next, consider \( II \), which can be written as

\[ II = \left( \frac{1}{N T^{3/2}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 (e_i^2 - w_i^{-1}) s_{it-1}^2 \right)^2 = \frac{I_1^2}{2II_2}. \]

By the same steps used for evaluating \( I_1 \),

\[ II_1 = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (e_i^2 - w_i^{-1}) s_{it-1}^2 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = II_1^p + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right), \]

implying that \( II_1 \) is asymptotically equivalent to \( II_1^p \) as \( N, T \rightarrow \infty \) with \( \frac{N}{T} \rightarrow \infty \). In order to compute the mean of this quantity note that by the unbiasedness of \( \hat{\sigma}^2 \), \( E(e_i^2 - \hat{\sigma}^2) = 0 \), which in turn implies that

\[ E(II_1^p) = \frac{1}{\sqrt{NT^{3/2}}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} E(e_i^2 - w_i^{-1})E(s_{it-1}^2) = 0. \]

33
The computation of the variance is simplified by rewriting $II^2_i$ in the following way:

$$II^2_i = \frac{1}{\sqrt{NT}^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\varepsilon^2_{it} - \bar{\varepsilon}_{it}^{-1})(s^2_{it-1} - \frac{1}{T} \sum_{k=p+1}^{T} s^2_{ik-1})$$

$$= \frac{1}{\sqrt{NT}^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\varepsilon^2_{it} - \bar{\varepsilon}_{it}^{-1})(s^2_{it-1} - \frac{1}{T} \sum_{k=p+1}^{T} s^2_{ik-1}) + O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right),$$

which uses deviations from means. Note that

$$\text{var}\left(\frac{1}{\sqrt{T}} \sum_{k=p+1}^{T} (\varepsilon^2_{ik} - 1)\right) = \frac{1}{T} \sum_{k=p+1}^{T} \sum_{i=1}^{T} E((\varepsilon^2_{ik} - 1)(\varepsilon^2_{i} - 1)) = \frac{1}{T} \sum_{k=p+1}^{T} E((\varepsilon^2_{ik} - 1)^2)$$

$$= \frac{1}{T} \sum_{k=p+1}^{T} (E(\varepsilon^2_{ik}) - 1) = \kappa_i - 1,$$

suggesting that

$$\frac{1}{\sqrt{T}} \sum_{k=p+1}^{T} (\varepsilon^2_{ik} - 1) \xrightarrow{w} \sqrt{\kappa_i - 1} V_i(r)$$

as $T \to \infty$, where $V_i(r)$ is a standard Brownian motion that is independent of $W_i(r)$, see Lemma A1 of McCabe and Tremayne (1995). It follows that

$$\frac{1}{T^{3/2}} \sum_{t=p+1}^{T} (\varepsilon^2_{it} - 1) (s^2_{it-1} - \frac{1}{T} \sum_{k=p+1}^{T} s^2_{ik-1}) \xrightarrow{w} \sqrt{\kappa_i - 1} \int_{0}^{1} \left(W_i(r)^2 - \int_{0}^{1} W_i(u)^2 du\right) dV_i(r),$$

from which we deduce

$$\text{var}(II^2_i) \to E\left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sqrt{\kappa_i - 1} \int_{0}^{1} \left(W_i(r)^2 - \int_{0}^{1} W_i(u)^2 du\right) dV_i(r)\right)^2\right]$$

$$= \frac{1}{N} \sum_{i=1}^{N} (\kappa_i - 1) \int_{0}^{1} E\left(\left(W_i(r)^2 - \int_{0}^{1} W_i(u)^2 du\right)^2\right) dr,$$

where

$$\int_{0}^{1} E\left(\left(W_i(r)^2 - \int_{0}^{1} W_i(u)^2 du\right)^2\right) dr = \int_{0}^{1} E(W_i(r)^4) dr - E\left(\int_{0}^{1} W_i(r)^2 dr\right)$$

$$= \int_{0}^{1} E(W_i(r)^4) dr - \int_{0}^{1} \int_{0}^{1} E(W_i(r)^2 W_i(u)^2) dr du.$$

By using the moments of Brownian motion,

$$\int_{0}^{1} E(W_i(r)^4) dr = 3 \int_{0}^{1} r^2 dr = 1,$$

$$\int_{0}^{1} \int_{0}^{1} E(W_i(r)^2 W_i(u)^2) dr du = \int_{0}^{1} \int_{0}^{1} (ru + 2 \min\{u^2, r^2\}) dr du = 2 \int_{0}^{1} \int_{0}^{1} (ru + 2u^2) dr du$$

$$= \frac{7}{3} \int_{0}^{1} r^3 dr = \frac{7}{12}.$$
and therefore
\[ \text{var}(II_1^o) \to \frac{5}{12} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\kappa_i - 1) = \frac{5}{12} (\kappa - 1) \]
as \(N, T \to \infty\). The results for the mean and variance of \(II_1^o\), together with Theorem 2 of Phillips and Moon (1999), yield \(II_1^o \to_d \sqrt{\frac{5}{12}} (\kappa - 1) Y\) as \(N, T \to \infty\) with \(\frac{N}{T} \to \infty\), where \(Y \sim N(0,1)\), implying \((II_1^o)^2 \to_d \frac{5}{24} (\kappa - 1) Y^2\).

As for \(II_2\), note that
\[
II_2 = \frac{1}{N T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 (2 \epsilon_{it}^2 - w_i^{-1}) s_{it-1}^4
\]

\[
= \frac{1}{N T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 (\epsilon_{it}^2 - w_i^{-1}) s_{it-1}^4 + \frac{1}{N T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 \epsilon_{it}^2 s_{it-1}^4
\]

\[
= \frac{1}{N T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_i^2 \epsilon_{it}^2 s_{it-1}^4 + O_p \left( \frac{1}{\sqrt{N T}} \right) = \frac{1}{N T^3} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \epsilon_{it}^2 s_{it-1}^4 + O_p \left( \frac{1}{\sqrt{T}} \right)
\]

where the last equality follows from the fact that \(II_1 = O_p(1)\). But
\[
\frac{1}{T^3} \sum_{t=p+1}^{T} \epsilon_{it}^2 s_{it-1}^4 \to_w \int_0^1 W_i(r)^4 dW_i(r)^2 = \int_0^1 W_i(r)^4 dr
\]
as \(T \to \infty\), suggesting that by Corollary 1 of Phillips and Moon (1999), as \(N, T \to \infty\)
\[
II_2^o \to_p \int_0^1 E(W_i(r)^4) dr = 3 \int_0^1 r^2 dr = 1.
\]

Consequently,
\[
II = \frac{II_1^2}{2II_2} = \frac{(II_1^o)^2}{2II_2^o} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) \to_d \frac{5}{24} (\kappa - 1) Y^2 \tag{A9}
\]
as \(N, T \to \infty\) with \(\frac{N}{T} \to \infty\).

It remains to show that \(X^2\) and \(Y^2\) are independent. Define
\[
X_N = \sqrt{\frac{5}{12}} \frac{\sum_{i=1}^{N} \int_0^1 W_i(r) dW_i(r)}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \int_0^1 W_i(r)^2 dr}}, \quad Y_N = \sqrt{\frac{12}{5}} \frac{\sum_{i=1}^{N} \int_0^1 W_i(r)^2 dV_i(r)}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \int_0^1 W_i(r)^4 dr}}
\]
such that \(X_N \to_d X\) and \(Y_N \to_d Y\) as \(N \to \infty\). Hence,
\[
ALM \to_d \lim_{N \to \infty} X_N^2 + \frac{5}{24} (\kappa - 1) \lim_{N \to \infty} Y_N^2
\]

35
as \( N, T \to \infty \) with \( \frac{N}{T} \to 0 \). By Corollary 5.3 or Park and Phillips (1988), \( X_1^2 \) and \( Y_1^2 \) are independent such that their sum is chi-squared distributed with two degrees of freedom. But \( W_i(r) \) and \( V_j(r) \) are independent for all \( i \) and \( j \), and therefore \( X_N^2 \) and \( Y_N^2 \) are also independent. The proof is completed by noting that the independence is preserved as \( N \to \infty \). ■

Proof of Theorem 2.

By combining (2) and (3) we get

\[
    w_{it} = \rho_i^t \phi_t(L)z_{ip} + \sum_{k=p+1}^t \rho_i^{t-k} \epsilon_{ik},
\]

which, via Taylor expansion and insertion of \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \), can be rewritten as

\[
    \frac{1}{\sqrt{T}} w_{it} = \frac{1}{\sqrt{T}} \sum_{k=p+1}^t \epsilon_{ik} + \frac{1}{\sqrt{T}} \phi_t(L)z_{ip} + \frac{c_i}{\sqrt{NT}} \left( \frac{t \phi_t(L)z_{ip} + \sum_{k=p+1}^t t-k \epsilon_{ik}}{T} \right) + o_p(1)
\]

\[
= \frac{1}{\sqrt{T}} \sum_{k=p+1}^t \epsilon_{ik} + o_p \left( \frac{1}{\sqrt{T}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) + o_p(1).
\]

(A10)

Hence, just as under \( H_0 \), if we assume that \( N, T \to \infty \), then \( \frac{1}{\sqrt{T}} \epsilon_{it} \to_w W_i(r) \). But we also have \( \Delta \epsilon_{it} = (\rho_i^t - 1) \epsilon_{it-1} + \epsilon_{it} \), from which it follows that

\[
(\Delta \epsilon_{it})^2 = ((\rho_i^t - 1) \epsilon_{it-1} + \epsilon_{it})^2 = (\rho_i^t - 1)^2 \epsilon_{it-1}^2 + 2(\rho_i^t - 1) \epsilon_{it-1} \epsilon_{it} + \epsilon_{it}^2
\]

\[
= \frac{c_i^2}{NT^2} \epsilon_{it-1}^2 + 2 \frac{c_i}{\sqrt{NT}} \epsilon_{it-1} \epsilon_{it} + \epsilon_{it}^2.
\]

Hence,

\[
\frac{1}{\sqrt{T}} \sum_{k=p+1}^t ((\Delta \epsilon_{ik})^2 - 1) = \frac{1}{\sqrt{T}} \sum_{k=p+1}^t (\epsilon_{ik}^2 - 1) + \frac{c_i}{\sqrt{NT^3/2}} \sum_{k=p+1}^t \epsilon_{ik}^2 + 2 \frac{c_i}{\sqrt{NT^3/2}} \sum_{k=p+1}^t \epsilon_{ik} \epsilon_{ik-1}
\]

\[
= \frac{1}{\sqrt{T}} \sum_{k=p+1}^t (\epsilon_{ik}^2 - 1) + o_p \left( \frac{1}{\sqrt{TN}} \right) \to_w \sqrt{k_i - 1} V_i(r) \quad \text{(A11)}
\]

as \( T \to \infty \), and by a similar calculation,

\[
\sigma_i^2 = \frac{1}{T-p} \sum_{t=p+1}^T (\Delta w_{it})^2 = \frac{c_i^2}{NT^3} \sum_{t=p+1}^T w_{it-1}^2 + 2 \frac{c_i}{\sqrt{NT^2}} \sum_{t=p+1}^T w_{it-1} \epsilon_{it} + \frac{1}{T} \sum_{t=p+1}^T \epsilon_{it}^2
\]

\[
= \frac{1}{T} \sum_{t=p+2}^T \epsilon_{it}^2 + o_p \left( \frac{1}{\sqrt{NT}} \right) = \sigma_i^2 + o_p \left( \frac{1}{\sqrt{T}} \right). \quad \text{(A12)}
\]

Equations (A10) to (A12) imply that the asymptotic results obtained for \((I_1^o)^2\), \((I_2^o)^2\), \((II_1^o)^2\) and \((II_2^o)^2\) under \( H_0 \) apply also under \( H_1 \).
Let us therefore decompose $ALM$ into $I = \frac{I^1}{T}$ and $II = \frac{I^2}{T}$, where $I^1$, $I_2$, $I^2_1$ and $I^2_2$ are the relevant numerator and denominator terms. We begin by considering $I$, where

\[ I_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{c}_{it} \hat{e}_{it-1} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} w_i \Delta \hat{c}_{it} \hat{e}_{it-1} \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta \hat{c}_{it} \hat{e}_{it-1} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \left( (\rho_i - 1)\hat{c}_{it-1}^2 + \hat{c}_{it-1} \hat{e}_{it} \right) + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = R_1 + I^1 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right), \]

where $I^1$ is as in the proof of Theorem 1, while

\[ E(R_1) = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+2}^{T} E(c_i \hat{e}_{it-1}^2) = \mu_c \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+2}^{T} E(\hat{e}_{it-1}^2) \rightarrow \frac{\mu_c}{4}, \]

which, via Corollary 1 of Phillips and Moon (1999), gives $R_1 \rightarrow p \frac{\mu_c}{2}$ as $N, T \rightarrow \infty$. Also, from the proof of Theorem 1, $I^1 \rightarrow_d \frac{1}{\sqrt{2}} X$, giving

\[ I^2_1 = (R_1 + I^1)^2 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = R_1^2 + 2R_1I^1 + (I^1)^2 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) \]

\[ \rightarrow_d \frac{\mu_c^2}{4} + \frac{\mu_c}{\sqrt{2}} X + \frac{1}{2} X^2. \]

But we also have that $I_2 \rightarrow_p \frac{1}{2}$ as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$, and so

\[ I \rightarrow_d \frac{\mu_c^2}{2} + \mu_c \sqrt{2} X + X^2. \quad \text{(A13)} \]

Next, consider $II$, where

\[ II_1 = \frac{1}{\sqrt{N} T^{3/2}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} ((\Delta \hat{c}_{it})^2 - 1)\hat{e}_{it-1}^2 \]

\[ = \frac{1}{\sqrt{N} T^{3/2}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \left( (\rho_i - 1)^2 \hat{c}_{it-1}^2 + 2(\rho_i - 1)\hat{c}_{it-1} \hat{e}_{it} - (\hat{e}_{it}^2 - w_i^{-1}) \hat{c}_{it-1} w_i^2 \hat{e}_{it-1}^2 \right) \]

\[ = R_1 + R_2 + II^1. \]

From the proof of Theorem 1 we know that $II^1 \rightarrow_d \frac{2}{T^2} (\kappa - 1) Y$ as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$. Also,

\[ R_1 = \frac{1}{N^{3/2} T^{7/2}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} c_i^2 w_i^2 \hat{e}_{it-1} = O_p \left( \frac{1}{\sqrt{NT}} \right), \]

\[ R_2 = \frac{2}{N T^{5/2}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} c_i w_i^2 \hat{e}_{it-1} \hat{e}_{it} = O_p \left( \frac{1}{\sqrt{NT}} \right). \]
Part \( II_2 \) can be written as

\[
II_2 = \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=p+2}^{T} (\Delta \tilde{x}_{it})^2 \tilde{x}_{it-1} + O_p \left( \frac{1}{\sqrt{NT}} \right)
\]

\[
= \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=p+2}^{T} (\rho_i - 1)^2 \tilde{x}_{it-1}^2 + 2(\rho_i - 1) \tilde{x}_{it}^2 \tilde{x}_{it-1} + O_p \left( \frac{1}{\sqrt{NT}} \right)
\]

\[
= R_1 + R_2 + II_2^0 + O_p \left( \frac{1}{\sqrt{NT}} \right),
\]

where \( II_2^0 \rightarrow_p 1 \), while

\[
R_1 = \frac{1}{N^2 T^3} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \tilde{c}_i \tilde{w}_i \tilde{e}_{it-1}^2 = O_p \left( \frac{1}{NT} \right),
\]

\[
R_2 = \frac{2}{N^3 T^7 / 2} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \tilde{c}_i \tilde{w}_i \tilde{e}_{it-1} \tilde{e}_{it} = O_p \left( \frac{1}{\sqrt{TN}} \right).
\]

Thus, by Taylor expansion,

\[
II = \frac{II_1}{II_2} = \frac{(II_2^0)}{II_2} + O_p \left( \frac{1}{\sqrt{NT}} \right) \rightarrow_d \frac{5}{24}(\kappa - 1)\gamma^2. \quad (A14)
\]

which, together with (A13), establishes the required result. \( \blacksquare \)

**Appendix C: Asymptotic properties of the feasible Lagrange multiplier statistic**

**Lemma C.1.** Under the conditions of Corollary 1 and in the model with a trend,

(a) \( \tilde{\omega}_{it} = \sum_{k=p+1}^{t} (\varepsilon_{ik} - \varepsilon_i) + O_p(1), \)

(b) \( \Delta \tilde{\omega}_{it} = \varepsilon_{it} + O_p \left( \frac{1}{\sqrt{T}} \right), \)

where \( \varepsilon_i = \frac{1}{T-p-1} \sum_{t=p+2}^{T} \varepsilon_{it}. \)

**Proof of Lemma C.1.**

We begin with (a). From (8) and Lemma A.1,

\[
\begin{align*}
\tilde{\omega}_{it} &= y_{it} - \hat{\Phi}_i\hat{y}_{it} - \hat{\mu}_i - \hat{\lambda}_i(t - p) \\
&= \sum_{k=p+1}^{t} \varepsilon_{ik} - (\hat{\Phi}_i - \Phi_i)' \hat{y}_{it} - (\hat{\mu}_i - \mu_i) - (\hat{\lambda}_i - \beta_i \phi_i(L))(t - p) \\
&= \sum_{k=p+1}^{t} \varepsilon_{ik} - (\hat{\Phi}_i - \Phi_i)' \hat{y}_{it} - (\hat{\mu}_i - \mu_i) - (\hat{\lambda}_i - \lambda_i)(t - p) - \beta_i \phi_i^*(1),
\end{align*}
\]
where the last equality uses the Beveridge–Nelson decomposition of $\phi_i(L)$ as $\phi_i(L) = \phi_i(1) + \phi_i^*(L)(1 - L)$.

Consider $\hat{\lambda}_i$. From (A2) we have that $\phi_i(L) \Delta y_{it} = \lambda_i + \epsilon_{it}$, or

$$\Delta y_{it} = \Phi_i' \Delta y_{it} + \lambda_i + \epsilon_{it}. \quad (A15)$$

Hence, $(\Delta y_{it} - \Delta \bar{y}_i) = \Phi_i'(\Delta y_{it} - \Delta \bar{y}_i) + \epsilon_{it} - \bar{\epsilon}_i$, which is a stationary regression with asymptotically exogenous regressors. It follows that as $T \to \infty$

$$\hat{\Phi}_i = \Phi_i + \left( \sum_{t=p+2}^{T} (\Delta y_{it} - \Delta \bar{y}_i)(\Delta y_{it} - \Delta \bar{y}_i)' \right)^{-1} \sum_{t=p+2}^{T} (\Delta y_{it} - \Delta \bar{y}_i)(\epsilon_{it} - \bar{\epsilon}_i).$$

Hence, since $\bar{\epsilon}_i = O_p(1/\sqrt{T})$, we have

$$\hat{\lambda}_i = \Delta \bar{y}_i - \Phi_i' \Delta \bar{y}_i = \lambda_i - (\Phi_i - \Phi_i)' \Delta \bar{y}_i + \bar{\epsilon}_i = \lambda_i + O_p\left(\frac{1}{\sqrt{T}}\right).$$

A similar calculation reveals that

$$\hat{\mu}_i = y_{ip+1} - \Phi_i' y_{ip} - \hat{\lambda}_i = \mu_i - (\Phi_i - \Phi_i)' y_{ip} - (\hat{\lambda}_i - \lambda_i) = \mu_i + O_p\left(\frac{1}{\sqrt{T}}\right).$$

By putting everything together,

$$\hat{\omega}_{it} = \sum_{k=p+1}^{t} (\epsilon_{ik} - \bar{\epsilon}_i) + O_p(1),$$

where we have made use of the fact that $\sum_{k=p+1}^{t} \epsilon_{ik} - (\hat{\lambda}_i - \lambda_i)(t - p) = \sum_{k=p+1}^{t} (\epsilon_{ik} - \bar{\epsilon}_i) + o_p(1)$. This establishes (a), and by similar arguments,

$$\Delta \hat{\omega}_{it} = \Delta y_{it} - \Phi_i' \Delta y_{it} - \hat{\lambda}_i = \epsilon_{it} - (\Phi_i - \Phi_i)' \Delta y_{it} - (\hat{\lambda}_i - \lambda_i) = \epsilon_{it} + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which establishes (b).

Note that if there is no trend in the model, then $(\hat{\lambda}_i - \lambda_i)(t - p)$ drops out in the equation for $\hat{\omega}_{it}$, and therefore so does $\bar{\epsilon}_i$. Hence, $\sum_{k=p+1}^{t} (\epsilon_{ik} - \bar{\epsilon}_i)$ reduces to $\sum_{k=p+1}^{t} \epsilon_{ik}$.

Note also that since the feasible maximum likelihood estimators converge to their unfeasible counterparts, and since $\epsilon_{ip+1}$ is zero when evaluated at the unfeasible estimators, observation $t = p + 1$ can be disregarded when forming the feasible Lagrange multiplier test statistic.
Proof of Corollary 1.

This proof follows by a simple adaptation of the proof of Theorem 2. We begin by writing the test statistic as
\[
FLM_1 = \frac{1}{\sqrt{NT}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} \Delta \hat{e}_{it} \hat{e}_{it-1} + 24 \left( \frac{1}{\sqrt{NT}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} \left( (\Delta \hat{e}_{it})^2 - 1 \right) \hat{e}_{it-1}^2 \right)
\]
\[
= \frac{I_1^2}{I_2} + \frac{24}{5(k-1)} II_1 II_2 = I + \frac{24}{5(k-1)} II
\]
with an obvious definition of $I_1, I_2, II_1$ and $II_2$.

Consider $I_1$. By Lemma C.1,
\[
\hat{\sigma}_i^2 = \frac{1}{T} \sum^{T}_{t=p+2} (\Delta \hat{\omega}_{it})^2 = \sigma_i^2 + O_p \left( \frac{1}{\sqrt{T}} \right) = \sigma_i^2 + O_p \left( \frac{1}{\sqrt{T}} \right), \tag{A16}
\]
which we can use to obtain
\[
I_1 = \frac{1}{\sqrt{NT}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} \Delta \hat{e}_{it} \hat{e}_{it-1} = \frac{1}{\sqrt{NT}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} \epsilon_{it} \epsilon_{it-1} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = I_1^o + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right)
\]
where the second equality uses the same trick as in the proof of Theorem 2, and where $I_1^o$ is the same as in that proof.

Similarly,
\[
I_2 = \frac{1}{\sqrt{NT^2}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} \epsilon_{it-1}^2 = \frac{1}{\sqrt{NT^2}} \sum^{N}_{i=1} \sum^{T}_{t=p+2} s_{it-1}^2 + O_p \left( \frac{1}{\sqrt{T}} \right) = I_2^o + O_p \left( \frac{1}{\sqrt{T}} \right),
\]
where $I_2^o$ is the same as before. It follows that
\[
I = \frac{(I_1)^2}{I_2^o} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) \overset{d}{\rightarrow} X^2
\]
as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$. But the same steps can be applied to show that $II \overset{d}{\rightarrow} \frac{5}{k^2} (\kappa - 1) X^2$. The proof is completed by noting that $k = \kappa + o_p(1)$.

Proof of Corollary 2.

We omit this proof in the paper. The required result is obtained by adapting the proof of Theorem 2 in the same way as the proof of Theorem 1 was adapted to establish Corollary 1.

Proof of Theorem 3.
Consider first the case when $H_0$ holds. Write

$$FLM_2 = \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta i_{it}^2 l_{it-1} + \frac{\sqrt{N}}{2} \right)^2 + \left( \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta i_{it}^2 l_{it-1} \right)^2 (k-1) \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \Delta i_{it}^2 l_{it-1}$$

$$= \frac{I_2^2}{I_2} + \frac{1}{k-1} II^2 = I + \frac{1}{k-1} II.$$

Let $g_{it} = \sum_{k=p+1}^{t} (\epsilon_{ik} - \bar{\epsilon}_i)$. By using Lemma C.1 and the technique of Theorem 2,

$$I = \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \epsilon_{it} g_{it-1} + \frac{\sqrt{N}}{2} \right)^2 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = \frac{(I_1^g)^2}{I_2^2} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right),$$

where $g_{it}^2 = (g_{it-1} + \Delta g_{it})^2 = g_{it-1}^2 + (\Delta g_{it})^2 + 2 g_{it-1} \Delta g_{it}$ with $g_{iT} = g_{i,p+1} = 0$, giving

$$\sum_{t=p+2}^{T} g_{it-1} \Delta g_{it} = \frac{1}{2} (g_{it}^2 - g_{i,p+1}^2) - \frac{1}{2} \sum_{t=p+2}^{T} (\Delta g_{it})^2 = -\frac{1}{2} \sum_{t=p+2}^{T} (\Delta g_{it})^2.$$

But $\Delta g_{it} = \epsilon_{it} + O_p (1/\sqrt{T})$ and hence

$$I_1^g = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \epsilon_{it} g_{it-1} + \frac{\sqrt{N}}{2} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} \Delta g_{it} g_{it-1} + \frac{\sqrt{N}}{2} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right),$$

$$= -\frac{1}{2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} ((\Delta g_{it})^2 - 1) + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right),$$

$$= -\frac{1}{2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} (\epsilon_{it}^2 - 1) + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right),$$

which uses the fact that $\frac{1}{T} \sum_{t=p+2}^{T} \epsilon_{it}^2 \rightarrow_p 1$ as $T \rightarrow \infty$.

Moreover, note that $\frac{1}{\sqrt{T}} g_{i,t-1} \rightarrow_w W_i(r) - r W_i(1)$ with $\frac{1}{T} \rightarrow r$ as $T \rightarrow \infty$, and so

$$\frac{1}{T^2} \sum_{t=p+2}^{T} E(g_{i,t-1}^2) \rightarrow \int_{0}^{1} E(W_i(r)^2 - 2rW_i(r)W_i(1) + r^2W_i(1)^2) dr = \frac{1}{6},$$

where we have used that $E(W_i(r)^2) = E(W_i(r)W_i(1)) = r$, and $E(W_i(1)^2) = 1$. Hence, by Corollary 1 of Phillips and Moon (1999), as $N, T \rightarrow \infty$

$$I_2^g = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+2}^{T} g_{i,t-1}^2 \rightarrow_p \frac{1}{6},$$

from which we deduce that

$$I = o_p(1) \quad (A17)$$
as \( N, T \to \infty \) with \( \frac{N}{T} \to 0 \).

Next, consider \( II \), which we write as

\[
II = \left( \frac{1}{\sqrt{NT^3/2}} \sum_{i=1}^N \sum_{t=p+1}^T (\varepsilon^2_t - w_t^{-1}) s^2_{it-1} \right)^2 + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = \frac{(II^o)^2}{II^o} + O_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right).
\]

Since \( \delta^2 \) is unbiased,

\[
II^o = \frac{1}{\sqrt{NT^3/2}} \sum_{i=1}^N \sum_{t=p+1}^T E(\varepsilon^2_t - w_t^{-1})E(s^2_{it-1}) = 0.
\]

The variance of \( II^o \) can be computed in the same way as in the proof of Theorem 1. We begin by rewriting \( II^o \) in terms of mean deviations, which gives

\[
II^o = \frac{1}{\sqrt{NT^3/2}} \sum_{i=1}^N \sum_{t=p+1}^T (\varepsilon^2_t - 1) \left( s^2_{it-1} - \frac{1}{T} \sum_{k=p+1}^T s^2_{ik-1} \right)
\]

as \( T \to \infty \), and therefore

\[
\text{var}(II^o) \to \frac{1}{N} \sum_{i=1}^N (\kappa_i - 1) \int_0^1 E \left[ \left( (W_i(r) - rW_i(1))^2 - \int_0^1 (W_i(u) - uW_i(1))^2 du \right) dV_i(r) \right] dr,
\]

where

\[
\int_0^1 \left( (W_i(r) - rW_i(1))^2 - \int_0^1 (W_i(u) - uW_i(1))^2 du \right) dr = \int_0^1 (W_i(r) - rW_i(1))^4 dr
\]

\[
- \left( \int_0^1 (W_i(r) - rW_i(1))^2 dr \right)^2.
\]

The expected value of the first term on the right-hand side is given by

\[
\int_0^1 E((W_i(r) - rW_i(1))^4) dr = \int_0^1 E(W_i(r)^4 - 4rW_i(1)^3W_i(r) + 6r^2W_i(1)^2W_i(r)^2
\]

\[
- 4r^3W_i(1)W_i(r)^3 + r^4W_i(r)^4) dr = \frac{1}{10}.
\]

where we have used that \( E(W_i(1)^3W_i(r)) = 3r^2, E(W_i(1)W_i(r)^3) = 3r, E(W_i(1)^2W_i(r)^2) = r + 2r^2 \) and \( E(W_i(1)^4) = 3 \). The second term can be expanded as

\[
\left( \int_0^1 (W_i(r) - rW_i(1))^2 dr \right)^2 = \int_0^1 \int_0^1 W_i(r)^2W_i(u)^2 drdu
\]

\[
- 4W_i(1) \int_0^1 \int_0^1 rW_i(r)W_i(u)^2 drdu + \frac{2}{3} W_i(1)^2 \int_0^1 W_i(r)^2 dr
\]

\[
+ 4W_i(1)^2 \int_0^1 \int_0^1 ruW_i(r)W_i(u) drdu - \frac{4}{3} W_i(1)^3 \int_0^1 rW_i(r) dr
\]

\[
+ \frac{1}{9} W_i(1)^4,
\]

42
where we know from before that the first term on the right-hand side has expectation $\frac{7}{12}$.

Moreover, since
\[
E(W_i(1)W_i(r)W_i(u)^2) = E(W_i(r)W_i(u)^3) + E((W_i(u) - W_i(r))W_i(r)W_i(u)^2)
\]
\[+ E((W_i(1) - W_i(u))W_i(r)W_i(u)^2)
\]
\[= E(W_i(u)^4) + E((W_i(r) - W_i(u))^2W_i(u)^2) = ru + 2u^2
\]
if $u < r$ and
\[E(W_i(1)W_i(r)W_i(u)^2) = E(W_i(r)^4) + 3E((W_i(u) - W_i(r))^2W_i(r)^2) = 3ru
\]
if $r < u$, we obtain
\[
\int_0^1 rE(W_i(1)W_i(r)W_i(u)^2)dr = \int_0^1 r \left( \int_0^r (ru + 2u^2)du + 3 \int_r^1 ru \right)dr = \frac{13}{36}
\]
and by a similar calculation, $\int_0^1 \int_0^1 ruE(W_i(1)^2W_i(r)W_i(u))drdu = \frac{16}{45}$. But we also have
\[
\int_0^1 E(W_i(1)^2W_i(r)^2)dr = \int_0^1 E((W_i(1) - W_i(r))^2W_i(r)^2 + W_i(r)^4)dr
\]
\[= \int_0^1 (r + 2r^2)dr = \frac{7}{6},
\]
\[
\int_0^1 rE(W_i(1)^3W_i(r))dr = \int_0^1 rE(W_i(1)^2W_i(r)^2)dr = \int_0^1 r(r + 2r^2)dr = 1,
\]
from which we obtain
\[
\int_0^1 E \left[ \left( (W_i(r) - rW_i(1))^2 - \int_0^1 (W_i(u) - uW_i(1))^2du \right)^2 \right] dr = \frac{1}{10} - \frac{1}{20} = \frac{1}{20}.
\]
It follows that as $N, T \to \infty$
\[
\text{var}(II_1^\circ) \to \frac{1}{20}(\kappa - 1),
\]
which, together with Theorem 2 in Phillips and Moon (1999), yields
\[II_1 \to_d \frac{1}{20}(\kappa - 1) Y^2
\]
as $N, T \to \infty$ with $\frac{N}{T} \to \infty$.

Also,
\[
E \left( \frac{1}{T^3} \sum_{t=p+2}^T e_{it}^2 s_{it-1}^4 \right) = \frac{1}{T^3} \sum_{t=p+2}^T E(e_{it}^2)E(s_{it-1}^4) = \frac{1}{T^3} \sum_{t=p+2}^T E(s_{it-1}^4)
\]
\[\to \int_0^1 E((W_i(r) - rW_i(1))^4)dr = \frac{1}{10}.
\]
Thus, since the conditions of Corollary 1 in Phillips and Moon (1999) are satisfied, $\mathscr{I}_2 \to_p \frac{1}{10}$ as $N, T \to \infty$ with $\mathcal{N}_T \to 0$, and so

$$\mathscr{I} \sim_d \frac{1}{2}(\kappa - 1)Y^2,$$  \hspace{1cm} (A18)

which establishes the required result under $H_0$.

In order to isolate the effect of the trend under $H_1$ note that from Lemma C.1,

$$\hat{\Delta \hat{w}}_{it} = \Delta y_{it} - \Phi'_p \Delta \bar{y}_{it} - \hat{\lambda}_i = \Delta y_{it} - \Phi'_p \Delta \bar{y}_{it} - \hat{\lambda}_i + \mathcal{O}_p \left( \frac{1}{\sqrt{T}} \right)$$

$$= \Delta y_{it} - \Delta \bar{y}_{it} - \Phi'_p (\Delta y_{it} - \Delta \bar{y}_{it}) + \mathcal{O}_p \left( \frac{1}{\sqrt{T}} \right) = \phi_i (L) (\Delta y_{it} - \Delta \bar{y}_{it}) + \mathcal{O}_p \left( \frac{1}{\sqrt{T}} \right).$$

Let $G_{it} = \sum_{k=p+1}^t s_{ik} - (t - p - 1) \bar{s}_i$, where $\bar{s}_i = \frac{1}{t - p - 1} \sum_{k=p+1}^{t-1} s_{ik}$. By using (1) and then (2),

$$\phi_i (L) (\Delta y_{it} - \Delta \bar{y}_{it}) = \phi_i (L) \left( \Delta z_{it} - \frac{1}{T - p - 1} \sum_{k=p+2}^T \Delta z_{ik} \right)$$

$$= (\rho_i - 1) \phi_i (L) \left( z_{it - 1} - \frac{1}{T - p - 1} \sum_{k=p+1}^{t-1} z_{ik} \right) + (\epsilon_{it} - \bar{\epsilon}_i),$$

$$= \sigma_i ((\rho_i - 1) (s_{it - 1} - \bar{s}_i) + \Delta g_{it}) + \mathcal{O}_p \left( \frac{1}{\sqrt{NT}} \right),$$

where the third equality uses that $\phi_i (L) z_{it} = \sum_{k=p+1}^t e_{ik} + \mathcal{O}_p (1)$. It follows that

$$\hat{\Delta \hat{w}}_{it} = \sigma_i ((\rho_i - 1) \Delta G_{it-1} + \Delta g_{it}) + \mathcal{O}_p \left( \frac{1}{\sqrt{T}} \right).$$  \hspace{1cm} (A19)

Similarly,

$$\hat{\hat{w}}_{it-1} = (\rho_i - 1) \phi_i (L) \left( \sum_{k=p+1}^{t-2} z_{ik} - \frac{t - p - 1}{T - p - 1} \sum_{k=p+1}^{t-1} z_{ik} \right) + \sum_{k=p+1}^{t-1} (\epsilon_{ik} - \bar{\epsilon}_i) + \mathcal{O}_p (1)$$

$$= \sigma_i ((\rho_i - 1) G_{it-2} + g_{it-1}) + \mathcal{O}_p (1).$$  \hspace{1cm} (A20)

These results, together with the consistency of $\hat{\sigma}^2_t$, imply

$$I_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=p+2}^T \hat{\Delta z}_{it} \hat{\epsilon}_{it-1} + \frac{\sqrt{N}}{2}$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=p+2}^T ((\rho_i - 1) \Delta G_{it-1} + \Delta g_{it}) ((\rho_i - 1) G_{it-2} + g_{it-1}) + \frac{\sqrt{N}}{2} + \mathcal{O}_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right)$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=p+2}^T ((\rho_i - 1)^2 \Delta G_{it-1} G_{it-2} + (\rho_i - 1) (\Delta g_{it} G_{it-2} + \Delta G_{it-1} g_{it-1})$$

$$+ \Delta g_{it} g_{it-1}) + \frac{\sqrt{N}}{2} + \mathcal{O}_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right) = R_1 + R_2 + I_1^* + \mathcal{O}_p \left( \frac{\sqrt{N}}{\sqrt{T}} \right).$$

44
where $I_1^* = o_p(1)$ as under $H_0$ and $R_1 = O_p(1/\sqrt{N})$.

Consider $R_2$. Note first that by Corollary 1 of Phillips and Moon (1999), as $N, T \to \infty$

$$R_2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} (\rho_i - 1)(\Delta g_{it} G_{it-2} + \Delta G_{it-1} g_{it-1})$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p+2}^{T} c_i(\Delta g_{it} G_{it-2} + \Delta G_{it-1} g_{it-2}) + \cdots$$

as for the second,

$$R_2 \to_p \mu_c \lim \frac{1}{T^2} \sum_{t=p+2}^{T} E(\Delta g_{it} G_{it-2} + \Delta G_{it-1} g_{it-1}).$$

Consider $E(\Delta g_{it} G_{it-2})$, which can be expanded as

$$E(\Delta g_{it} G_{it-2}) = E \left[ (\varepsilon_{it} - \bar{\varepsilon}_i) \left( \sum_{k=p+1}^{t-2} s_{ik} - (t - p - 1)\bar{s}_i \right) \right] = E \left( \varepsilon_{it} \sum_{k=p+1}^{t-2} s_{ik} \right) - (t - p - 1)E(\varepsilon_i \bar{s}_i) - E \left( \bar{\varepsilon}_i \sum_{k=p+1}^{t-2} s_{ik} \right) + (t - p - 1)E(\varepsilon_i \bar{s}_i),$$

where the first term on the right-hand side is zero, while as for the second,

$$E(\varepsilon_i \bar{s}_i) = \frac{1}{T - p - 1}E \left( \varepsilon_i \sum_{k=i}^{T-1} s_{ik} \right) = \frac{T - t}{T - p - 1}.$$

Similarly,

$$E \left( \bar{\varepsilon}_i \sum_{k=p+1}^{t-2} s_{ik} \right) = \frac{1}{T - p - 1}E \left( \sum_{i=p+2}^{T-1} \varepsilon_{it} \sum_{k=p+1}^{t-2} s_{ik} \right) = \frac{(t - p - 2)(t - p - 3)}{2(T - p - 1)},$$

$$E(\varepsilon_i \bar{s}_i) = \frac{1}{(T - p - 1)^2}E \left( \sum_{i=p+2}^{T-1} \varepsilon_{it} \sum_{k=p+2}^{T-1} s_{ik} \right) = \frac{T - p - 2}{2(T - p - 1)},$$

which yields

$$E(\Delta g_{it} G_{it-2}) = -\frac{1}{2(T - p - 1)} \left[ (t - p - 2)(t - p - 3) + 2(T - t)(t - p - 1) \right] - (T - p - 2)(t - p - 1).$$

(A22)

Next, consider $E(\Delta G_{it-1} g_{it-1})$. It holds that

$$E(\Delta G_{it-1} g_{it-1}) = E((s_{it-1} - \bar{s}_i)(s_{it-1} - (t - p - 1)\bar{s}_i)) = E(s_{it-1}^2) - (t - p - 1)E(s_{it-1}\bar{s}_i) - E(s_{it-1}\bar{s}_i) + (t - p - 1)E(\varepsilon_i \bar{s}_i),$$

where $E(s_{it-1}^2) = t - p - 1$, implying

$$E(s_{it-1}\bar{s}_i) = \frac{1}{T - p - 1}E \left( s_{it-1} \sum_{k=p+1}^{T-1} \varepsilon_{ik} \right) = \frac{1}{T - p - 1}E(s_{it-1}^2) = \frac{t - p - 1}{T - p - 1}. $$

45
Also,
\[
E(s_{it-1}s_{i}) = \frac{1}{T-p-1} E \left[ s_{it-1} \left( \sum_{k=p+1}^{t-1} s_{ik} + \sum_{k=t}^{T-1} s_{ik} \right) \right] \\
= \frac{t-p-1}{2(T-p-1)} \left( (t-p-2) + 2(T-t) \right),
\]
\[
E(\varepsilon_is_{i}) = \frac{1}{(T-p-1)^2} E \left( \sum_{k=p+1}^{T-1} \varepsilon_{ik} \sum_{k=p+1}^{T-1} s_{ik} \right) = \frac{T-p-2}{2(T-p-1)},
\]
from which we deduce that
\[
E(\Delta G_{it}s_{i-1}) = \frac{(t-p-1)(T-t)}{2(T-p-1)}.
\]

Equations (A21) to (A23) imply
\[
R_2 \to_p \mu_c \lim_{T \to \infty} \frac{1}{T^2} \sum_{t=p+2}^{T} \frac{t-p-2}{T-p-1} = O \left( \frac{1}{T} \right).
\]

Hence, \( I_1 = o_p(1) \) as \( N, T \to \infty \) with \( \frac{N}{T} \to 0 \), and we already know from before that \( I_2 \to_p \frac{1}{2} \).

Therefore, \( I = o_p(1) \). But it also holds that \( II \Rightarrow \frac{1}{2}(\kappa - 1) Y^2 \), and so the proof is complete. \( \blacksquare \)
Table 1: Size and size-adjusted power at the 5% level for model 1.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$\phi = 0$</th>
<th>$\phi = 0.5$</th>
<th>$\phi = -0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$FLM_1$ $t_z^*$ $Z_{bar}$</td>
<td>$FLM_1$ $t_z^*$ $Z_{bar}$</td>
<td>$FLM_1$ $t_z^*$ $Z_{bar}$</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>8.1  14.7 7.5</td>
<td>7.5  21.9 8.0</td>
<td>8.3  7.3 7.3</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>10.0 21.2 8.8</td>
<td>8.9  33.3 8.9</td>
<td>10.3 9.0 8.6</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>6.5  10.3 6.2</td>
<td>6.0  15.2 6.4</td>
<td>6.7  6.1 6.0</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>6.9  13.3 6.0</td>
<td>6.8  21.7 6.4</td>
<td>7.4  6.3 5.9</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>5.9  5.3 4.8</td>
<td>6.0  8.3 5.1</td>
<td>6.1  3.5 4.5</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>6.2  6.7 5.4</td>
<td>5.9  11.5 5.5</td>
<td>6.4  3.6 5.4</td>
</tr>
</tbody>
</table>

Case 1: $\rho_i = 1$ for all $i$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$\phi = 0$</th>
<th>$\phi = 0.5$</th>
<th>$\phi = -0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10</td>
<td>38.2 9.2 33.7</td>
<td>28.1 8.4 30.3</td>
<td>42.7 6.9 35.2</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>43.9 9.6 35.3</td>
<td>32.5 7.9 33.6</td>
<td>48.2 7.2 36.1</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>50.3 6.1 37.0</td>
<td>43.5 5.4 34.8</td>
<td>52.7 4.7 38.2</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>57.7 5.6 39.4</td>
<td>50.1 5.5 36.3</td>
<td>59.7 4.3 40.5</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>56.3 4.0 41.8</td>
<td>52.6 4.1 38.6</td>
<td>57.9 3.2 41.8</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>67.7 2.9 38.8</td>
<td>66.0 3.4 38.5</td>
<td>68.0 2.2 39.8</td>
</tr>
</tbody>
</table>

Case 2: $\rho_i = 1 + \frac{c_i}{\sqrt{NT}}$ with $c_i = -10$ for all $i$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$\phi = 0$</th>
<th>$\phi = 0.5$</th>
<th>$\phi = -0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10</td>
<td>30.8 10.2 34.1</td>
<td>23.5 8.8 29.3</td>
<td>33.5 8.2 35.7</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>34.9 10.1 35.3</td>
<td>27.1 8.7 32.5</td>
<td>38.2 8.1 36.5</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>40.4 7.1 37.2</td>
<td>35.0 7.1 34.4</td>
<td>41.2 5.8 37.9</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>46.9 6.5 40.1</td>
<td>41.7 6.1 36.5</td>
<td>48.9 5.4 41.5</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>46.2 5.2 42.1</td>
<td>43.4 5.1 39.2</td>
<td>47.1 4.3 42.0</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>56.2 3.9 39.9</td>
<td>54.7 4.1 38.9</td>
<td>55.7 3.0 41.2</td>
</tr>
</tbody>
</table>

Case 3: $\rho_i = 1 + \frac{c_i}{\sqrt{NT}}$ with $c_i \sim U(-20, 0)$

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$\phi = 0$</th>
<th>$\phi = 0.5$</th>
<th>$\phi = -0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10</td>
<td>36.6 0.8 2.9</td>
<td>49.3 0.5 2.4</td>
<td>21.7 0.9 3.0</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>17.5 0.0 0.3</td>
<td>32.9 0.0 0.3</td>
<td>7.4 0.0 0.2</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>43.0 0.7 3.2</td>
<td>35.5 0.6 3.1</td>
<td>43.4 0.6 3.3</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>36.5 0.1 0.5</td>
<td>24.5 0.1 0.6</td>
<td>57.4 0.1 0.5</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>61.9 0.3 4.2</td>
<td>39.8 0.3 4.0</td>
<td>64.8 0.3 4.4</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>72.3 0.0 0.9</td>
<td>37.3 0.0 1.0</td>
<td>77.5 0.0 1.0</td>
</tr>
</tbody>
</table>

Case 4: $\rho_i = 1 + \frac{c_i}{\sqrt{NT}}$ with $c_i \sim U(-40, 20)$

continued overleaf
Table 1: Continued.

<table>
<thead>
<tr>
<th></th>
<th>( \phi = 0 )</th>
<th>( \phi = 0.5 )</th>
<th>( \phi = -0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T )</td>
<td>( N )</td>
<td>( FLM_1 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>26.3</td>
<td>7.2</td>
</tr>
<tr>
<td>20</td>
<td>52.7</td>
<td>11.0</td>
<td>41.6</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>83.9</td>
<td>13.4</td>
</tr>
<tr>
<td>20</td>
<td>99.5</td>
<td>26.1</td>
<td>96.7</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>99.9</td>
<td>50.2</td>
</tr>
<tr>
<td>20</td>
<td>100.0</td>
<td>84.7</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>22.7</td>
<td>7.9</td>
</tr>
<tr>
<td>20</td>
<td>41.2</td>
<td>11.6</td>
<td>41.5</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>62.8</td>
<td>14.8</td>
</tr>
<tr>
<td>20</td>
<td>88.0</td>
<td>25.8</td>
<td>93.4</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>90.9</td>
<td>40.8</td>
</tr>
<tr>
<td>20</td>
<td>98.7</td>
<td>64.6</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>19.7</td>
<td>2.0</td>
</tr>
<tr>
<td>20</td>
<td>17.2</td>
<td>0.4</td>
<td>2.2</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>48.0</td>
<td>3.2</td>
</tr>
<tr>
<td>20</td>
<td>43.4</td>
<td>0.5</td>
<td>4.8</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>77.2</td>
<td>5.3</td>
</tr>
<tr>
<td>20</td>
<td>71.6</td>
<td>0.7</td>
<td>5.9</td>
</tr>
</tbody>
</table>

Notes: The parameter \( \phi \) refers to the autoregressive coefficient, while \( t^*_d \) and \( Z_{\text{bar}} \) refer to the tests of Levin et al. (2002) and Im et al. (2003), respectively.
Table 2: Size and size-adjusted power at the 5% level for model 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( \phi = 0 )</th>
<th>( \phi = 0.5 )</th>
<th>( \phi = -0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FLM</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( T )</td>
<td>( N )</td>
<td>( t_{\delta}^* )</td>
<td>( Z_{\text{bar}} )</td>
<td>( t_{\delta}^* )</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>6.9</td>
<td>20.0</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>8.2</td>
<td>31.7</td>
<td>13.9</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>6.0</td>
<td>14.9</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>7.5</td>
<td>20.4</td>
<td>9.6</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>6.0</td>
<td>2.7</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6.5</td>
<td>2.6</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Case 1: \( \rho_i = 1 \) for all \( i \)

Case 2: \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i = -10 \) for all \( i \)

Case 3: \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i \sim U(-20, 0) \)

Case 4: \( \rho_i = 1 + \frac{c_i}{\sqrt{NT}} \) with \( c_i \sim U(-40, 20) \)

\[ \text{Continued overleaf} \]
Table 2: Continued.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\rho_i = 1 + c_i$ with $c_i = -0.05$ for all $i$</th>
<th>$\rho_i = 1 + c_i$ with $c_i \sim U(-0.1, 0)$</th>
<th>$\rho_i = 1 + c_i$ with $c_i \sim U(-0.15, 0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T$</td>
<td>$N$</td>
<td>$FLM_1$</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>5.7</td>
<td>7.6</td>
</tr>
<tr>
<td>20</td>
<td>7.1</td>
<td>9.1</td>
<td>12.5</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>8.4</td>
<td>15.3</td>
</tr>
<tr>
<td>20</td>
<td>9.6</td>
<td>20.6</td>
<td>47.1</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>11.6</td>
<td>45.9</td>
</tr>
<tr>
<td>20</td>
<td>17.1</td>
<td>74.5</td>
<td>99.7</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>5.9</td>
<td>7.8</td>
</tr>
<tr>
<td>20</td>
<td>7.7</td>
<td>9.8</td>
<td>14.5</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>7.6</td>
<td>17.7</td>
</tr>
<tr>
<td>20</td>
<td>8.7</td>
<td>23.8</td>
<td>51.5</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>8.8</td>
<td>44.6</td>
</tr>
<tr>
<td>20</td>
<td>12.0</td>
<td>67.6</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Notes: See Table 1.
### Table 3: Empirical results from the feasible Lagrange multiplier test.

<table>
<thead>
<tr>
<th>Factor treatment</th>
<th>Prices</th>
<th>Model</th>
<th>Test</th>
<th>p-value</th>
<th>Income</th>
<th>Test</th>
<th>p-value</th>
<th>Price-to-income</th>
<th>Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No factors allowed</td>
<td>1</td>
<td>42.58</td>
<td>0.00</td>
<td>154.85</td>
<td>0.00</td>
<td>170.92</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>13.66</td>
<td>0.00</td>
<td>36.62</td>
<td>0.00</td>
<td>17.36</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal components</td>
<td>1</td>
<td>10.05</td>
<td>0.01</td>
<td>3.48</td>
<td>0.18</td>
<td>4.39</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.20</td>
<td>0.07</td>
<td>0.27</td>
<td>0.61</td>
<td>5.20</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No factors allowed</td>
<td>1</td>
<td>91.89</td>
<td>0.00</td>
<td>248.79</td>
<td>0.00</td>
<td>100.61</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No factors allowed</td>
<td>2</td>
<td>20.40</td>
<td>0.00</td>
<td>71.33</td>
<td>0.00</td>
<td>25.31</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal components</td>
<td>1</td>
<td>36.62</td>
<td>0.00</td>
<td>29.98</td>
<td>0.00</td>
<td>21.37</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal components</td>
<td>2</td>
<td>7.50</td>
<td>0.01</td>
<td>8.78</td>
<td>0.00</td>
<td>1.63</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The principal components method was implemented with the number of factors estimated using the IC2 criterion of Bai and Ng (2002). The order of the lag augmentation in the tests was estimated by using the Schwarz Bayesian information criterion.